REPORT RESUMES

ED 010 872 AL 000 023

LINGUISTIC STUDIES FOR CHINESE TO ENGLISH MACHINE TRANSLATION.

BY- LEE, THERESA AND OTHERS
ITEK CORP., LEXINGTON, MASS.
ROME AIR DEVELOPMENT CENTER, GRIFFISS AFB, N.Y.
REPORT NUMBER RADC-TR-65-433 PUB DATE JUN 66
EDRS PRICE MF-$0.27 HC-$6.84 171P.

DESCRIPTORS- *CHINESE, *ENGLISH, *COMPUTATIONAL LINGUISTICS, *MACHINE TRANSLATION, MORPHOLOGY (LANGUAGES), SYNTAX, DATA PROCESSING, CONTEXTUAL ASSOCIATIVE METHOD, LEXINGTON

A LINGUISTIC PROCESSING SYSTEM, WHICH HAS BEEN DEVELOPED FOR CHINESE TO ENGLISH MACHINE TRANSLATION, UTILIZING THE CONTEXTUAL ASSOCIATIVE METHOD WAS DESCRIBED. THE REPORT INCLUDES (1) AN EXPLANATION OF THE LINGUISTIC PROCESSING SYSTEM, (2) MORPHOLOGICAL AND SYNTACTIC ANALYSES, AND (3) AN ANALYSIS OF ENGLISH INFLECTION. MACHINE APPLICABILITY WAS STRESSED THROUGHOUT, AND TESTING OF LINGUISTIC RULES BY MACHINE WAS INITIATED AND PROVED VALUABLE. THE CHINESE INPUT WAS PROCESSED BY CHICODER, A DEVICE USED TO ENCODE CHINESE CHARACTERS. (KL)

ED010872

AL 000 023

RADC-TR-65-443
FINAL REPORT

LINGUISTIC STUDIES FOR CHINESE TO ENGLISH MACHINE TRANSLATION

Mrs. Theresa L[illegible] et al

TECHNICAL REPORT NO[illegible]ADC-TR-65-443

June 1966

This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of RADC (EMLI), GAFB, N.Y. 13440.

Rome Air Development Center
Research and Technology Division
Air Force Systems Command
Griffiss Air Force Base, New York

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related government procurement operation, the government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded, by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacturer, use, or sell any patented invention that may in any way be related thereto.

[redacted]

Do not return this copy. Retain or destroy.

LINGUISTIC STUDIES FOR CHINESE TO ENGLISH MACHINE TRANSLATION

Mrs. Theresa Lee, et al

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

This document is subject to special
export controls and each transmittal
to foreign governments or foreign
nationals may be made only with
prior approval of RADC (EMLI),
GAFB, N.Y. 13440.

AFLC, GAFB, N.Y., 8 Aug 66-164

FOREWORD

This report was prepared by Itek Corporation, 10 Maguire Road, Lexington 73, Massachusetts; under Contract Number AF30(602)-3494 and Project Number 4599. The RADC project engineer is Zbigniew L. Pankowicz, EMIIH.

This is a final report and covers the period of work from 1 July 1964 to 30 June 1965.

This report is intended to document research efforts in linguistic studies for Chinese to English machine translation. The objectives to be reached under the contract are as follows:

1. Morphological and syntactical analysis of modern Chinese for machine translation applications.
2. Compilation of 15,000 Chinese input entries and 15,000 English output entries on magnetic tape.
3. Compilation of linguistic rules in symbolic notation.

This work covers a 1-year period of main emphasis on linguistic research in the area of morphological and syntactical problems in modern Chinese. As part of a company sponsored, independent research effort, programming testing efforts were incorporated in the second half of the year to test the validity of the linguistic rules and the practicability of the operations for machine application. The company sponsored effort is detailed in Section 2.5 and in Appendix B of this report, to fulfill the requirements of part VI, paragraph e-3, of the contract schedule.

The contract research efforts have resulted in the following:

1. A linguistic machine translation system from the Chicoder input of the source language to the linguistic rule operations to the output of the target language.
2. A magnetic tape and a list of 15,000 Chinese input entries and 30,000 English output entries plus input and output entries of specialized vocabulary of general function words.
3. A list of linguistic rules, in symbolic notation, that is divided into sections according to operations specified in this report (the linguistic rules in symbolic notation are contained in a separate report, "Linguistic Rules in Symbolic Notation," October 18, 1965, Itek Corporation).

FOREWORD (Cont)

4. Details of programming testing of the linguistic rules and of testing of operations for these rules.

The research efforts were made by Itek linguistic research personnel under the direction of Mrs. Theresa C. Lee of the Chinese Programs Group. Additional authors of this report are: H. T. Wang, S. C. Yang, and E. L. Farmer.

Special assistance was rendered by Mr. C. R. Burgess of the Computer Sciences Department. The late Dr. Jennings Wong contributed significantly to the research efforts of this project in the source and target language analysis.

Special acknowledgement is noted here for the assistance rendered by Mr. Zbigniew L. Pankowicz and Mr. Wing Y. Hoo, the technical monitors of this project at RADC during this year of study.

Release of subject report to the general public is prohibited by the Strategic Trade Control Program, Mutual Defense Assistance Control List (revised 6 January 1965), published by the Department of State.

This technical report has been reviewed and is approved.

Approved:
FRANK J. TOMAINI
Chief, Info Prcs Branch
Intel and Info Processing Div.

Approved:
fr Howard Davis
ROBERT J. QUINN, JR., Colonel, USAF
Chief, Intel and Info Processing Div

FOR THE COMMANDER:
IRVING J. GABELMAN
Chief, Advanced Studies Group

ABSTRACT

The linguistic study to increase the capabilities of a developmental Chinese to English machine translation has led to the design of a basic linguistic processing system utilizing the Contextual Associative Method (CAM). This technique allows machine translation through the use of programmed contextual operations.

The results of the research effort are presented in this report and include: (1) explanation of the linguistic processing system, (2) morphological and syntactical analyses, and (3) English inflection analysis for Chinese to English machine translation. Illustrations showing step by step linguistic processing are included in this report. Recommendations are presented for refinement and further development of the basic linguistic analysis to further the goal of performing Chinese to English machine translation. Three appendices are included: an explanation of symbols for linguistic rules (Appendix A), listings of computer experimentation (Appendix B), and listings of verb components in English output (Appendix C).

PRECEEDING PAGE BLANK NOT FILMED

TABLE OF CONTENTS

1. INTRODUCTION 1-1

2. LINGUISTIC PROCESSING SYSTEM FOR CHINESE TO ENGLISH MACHINE TRANSLATION 2-1

2.1 Description of Chicoder and Its Function in Machine Translation . . 2-1
2.2 Dictionary Entry Format and Its Relationship to the Linguistic Processing System 2-3
2.3 Linguistic Rule Entry Format and Its Relationship to the Linguistic Processing System 2-5
2.4 Linguistic Processing System 2-6
2.5 Computer Experimentation 2-12

3. MORPHOLOGICAL ANALYSIS 3-1

3.1 Morphology Tags and Explanations 3-1
3.2 Nominals 3-13
3.3 Adjectival Adverbials 3-19
3.4 Verbals 3-23
3.5 Arithmates 3-31
3.6 Synonomes 3-31
3.7 Numerals 3-31
3.8 Conjunctions 3-32
3.9 Auxiliaries 3-32
3.10 Con-Adjectival-Adverbials 3-32
3.11 Punctuation 3-35
3.12 Special Words - H Class 3-37
3.13 Connominals 3-37
3.14 Converbals 3-42
3.15 Collocatives 3-42

4. SYNTACTICAL ANALYSIS 4-1

4.1 Noun Phrases 4-1
4.2 Verb Phrases 4-9
4.3 Collocative Words and Structures 4-14
4.4 Connominal Phrases 4-19
4.5 Noun Phrase-Relative Clauses 4-24
4.6 Verb Linking 4-31
4.7 Illustrations of Linguistic Processing on a Sentence Level 4-32

PRECEEDING PAGE BLANK NOT FILMED

TABLE OF CONTENTS (Cont)

5. ENGLISH WORD INFLECTIONS 5-1

5.1 Noun Inflections 5-1
5.2 Adjective and Adverb Auxiliaries and Inflections 5-5
5.3 Verbs 5-8
5.4 Noun, Adjective-Adverb, and Verb Table Designators 5-14

6. CONCLUSIONS AND RECOMMENDATIONS 6-1

6.1 Morphology 6-2
6.2 Syntax 6-3

7. PROJECT PERSONNEL 7-1

8. BIBLIOGRAPHY 8-1

APPENDIX A - EXPLANATION OF SYMBOLS FOR LINGUISTIC RULES . . . A-1

APPENDIX B - LISTINGS OF COMPUTER EXPERIMENTATION B-1

APPENDIX C - VERB COMPONENTS IN ENGLISH OUTPUTS C-1

LIST OF ILLUSTRATIONS

2-1 Sample Dictionary Entry Card 2-4

LIST OF TABLES

2-1 Linguistic Passes 2-7/8

3-1 General Table of Chinese Morphological Tags 3-2

3-2 Specific Table of the Nominal Class Tags 3-3/4

3-3 Specific Table of the Adjectival-Adverbial Class Tags 3-5

3-4 Specific Table of the Verbal Class Tags 3-7/8

3-5 Specific Tables for the Arithmate, Synonome, Numeral, Conjunction, Auxiliary, Con-Adjectival-Adverbial, Punctuation, and Special Class Tags 3-9

3-6 Specific Table for the Connominal Class Tags 3-10

3-7 Specific Table for the Converbal Class Tags 3-11

3-8 Specific Table for the Collocative Class Tags 3-11

3-9 Symbols, Terminology, and General Definition for Major Chinese Word Classes 3-12

4-1 Verb Linking Sentence Patterns 4-33

4-2 Sample Sentence I 4-35/36

4-3 Rules Used in Sample Sentence I 4-37

4-4 Sample Sentence II 4-39/40

4-5 Rules Used in Sample Sentence II 4-41

5-1 Singular and Plural Table Designators 5-3

5-2 Plural Position Table Designators 5-4

5-3 Possessive Table Designators 5-4

5-4 Adjective Table Designators 5-6

5-5 Adverb Table Designators 5-7

5-6 The Six Basic Verb Stem Forms 5-9

LIST OF TABLES (Cont)

5-7 The Most Common Forms of Stem Change for English Verbs 5-9
5-8 Forms of Stem Changes for Weak or Regular Verbs 5-10
5-9 Examples of Verb Stems and Ending Formations 5-12
5-10 Nine Forms of the Verb "To Be" 5-12
5-11 Auxiliary Verbs That Can Occur in English Output 5-13
5-12 Negative Auxiliary Verbs 5-15
5-13 Positive and Negative Infinitive Auxiliaries 5-16
5-14 Positive and Negative Postmodal Auxiliaries 5-17
5-15 Table of Noun, Verb, and Adjective-Adverb Table Designators . . . 5-18

EVALUATION

The machine translation R&D effort described in subject TR encompasses the following activities: (1) linguistic analysis of source language; (2) compilation of 15,000 bilingual entries covering the field of political science; and (3) formulation of linguistic rules for processing of source language data. The machine translation system thus constructed uses the Chicoder as an input device and the photostore type equipment as a central processor. Preliminary results of this R&D effort have been programmed and partly tested on the PDP-1 System.

The linguistic research described in subject PR is oriented toward obtaining early practical results for application to machine translation. The authors recommend further linguistic studies in morphology and syntax as a means of improving the English output.

Zbigniew L. Pankowicz

ZBIGNIEW L. PANKOWICZ
Textual Data Handling Section
Information Processing Branch

1. INTRODUCTION

The research efforts for the past year have been devoted to the linguistic studies of Chinese to English machine translation. In this course of study, a system has been developed for linguistic processing from the input of Chinese text to the output of English. For this process of development, the following major goals have been set:

1. To orient the linguistic research toward the practicability of machine application
2. To develop a practicable system for transferring the source language, Chinese, to the target language, English, by analyzing the source language with the use of modern Chinese political texts, which cover many morphological and syntactical functions of the source language
3. To formulate a grammatical tag system for Chinese morphology and to formulate a linguistic rule operation system for Chinese syntax
4. To formulate an English output system to be compatible with input
5. To provide, on magnetic tape, 15,000 lexicographic entries in the field of political science with appropriate grammatic tags and to add general vocabulary input and output entries, which are utilized in the linguistic processing system
6. To determine a programming system for the processing from input to output in testing the linguistic rules.

Our linguistic analysis approach is an attempt to bring the source language and the target language closer to each other. The input of the source language and the quality of translation derived from the linguistic analyses are therefore considered from every aspect of the problems involved and their resolutions.

Political science texts such as "Ren Min Ri Bao"[23]* and "Hong Qi"[14] were studied continuously during the year for a constant sampling of texts to modify or enhance the linguistic study, since modification of language structures is constant and the best source of research is the actual writing in texts. Machine oriented linguistic analysis texts and general linguistic texts, as listed in Section 8, were studied for ideas that are applicable to the present research.

The first task was to determine the major morphological classes through analysis of Chinese words. Each major morphological class was given a series of subtags that define the grammatic subtlety of the morphological class. The morphological grammatic tags were modified or redefined as research progressed to the syntactic level.

The second task, which was concurrent to the first, was the syntactical analysis. On this level the words were grouped into major phrases and attributive clauses and phrases. These phrases and clauses were analyzed with consideration given to the relative importance of their roles in the sentence. Within this task the problems of linking major phrases in a sentence and of sentence patterns were also considered.

The third task was to formulate and analyze English inflection tables for proper translation output. The major word classes under consideration for this task were the nominals, verbals, adjectivals, and adverbials.

During the latter half of this study, applicability of the linguistic analysis to machine operations was emphasized, and this recognition of eventual machine implementation continually influenced the linguistic processing system composition throughout the rest of the period.

When the basic theories of morphological and syntactical analysis were established, the production of lexicographic entries and linguistic rules began. It is to be emphasized that all categorizing of linguistic operations and lexicographic work is based on study of the source and target languages and machine applicability.

This project is an attempt to bring the source language closer to the target language morphologically, syntactically, and in the relationship of sentence structures to patterns. The Contextual Associative Method first looks up the process sentence morphologically. The major and attributive structures are then segmented for processing within the structures, and the major structures are linked for correct processing of output.

---

* References may be found in the bibliography (Section 8) of this report.

2. LINGUISTIC PROCESSING SYSTEM FOR CHINESE TO ENGLISH MACHINE TRANSLATION

The linguistic processing of Chinese to English machine translation is divided into the following steps:

1. Operation of the Chicoder to generate input tape for the language processor
2. Lookup of Chinese grammatic tags from information on input tape and in dictionary entry tables on magnetic tape or disc
3. Performance of linguistic sentence analysis through the use of linguistic rule tables, both programming and lookup techniques to be used in deriving English grammatic tags
4. Lookup of word stem of the appropriate English translation from English grammatic tags
5. Lookup of appropriate English forms for the word stem.

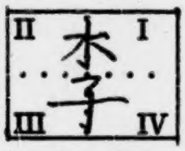
2.1 DESCRIPTION OF CHICODER AND ITS FUNCTION IN MACHINE TRANSLATION

The Chicoder is a device, designed to encode Chinese characters, that was completed under an Air Force contract awarded by RADC. It has the same number of keys as the English typewriter; all codes therefore refer to the alphabet letters or Arabic numerals on the English typewriter keys. The Chicoder has a vocabulary of 10,518 Chinese characters, and it is designed so that 90 percent of the common Chinese characters can be encoded in three strokes.

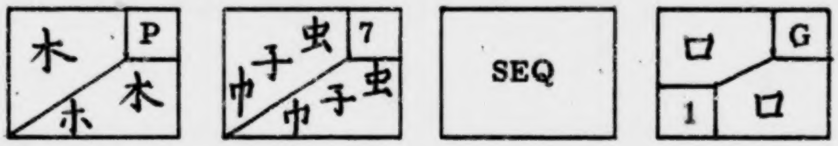
The Chicoder is designed to operate in two modes, English and Chinese. In the English mode, the keyboard functions like the regular English typewriter in the lower case. In the Chinese mode Chinese characters may be encoded, and the keyboard functions like the English typewriter in the upper case. Punctuation marks and mathematical symbols are in the Chicoder code (1v).

Two codes are typed to position a character, and a line of five Chinese characters is displayed on the screen. The character is selected by typing one of the position keys designated on the lower left corner of the a, s, d, f, and g keys. If the character is not on the first line, the sequencing (SEQ) key is depressed for the second line. If it is not there, SEQ is depressed for the third line, and so on.

The encoding scheme looks at the whole character as a unit square divided into quadrants, i.e., I, II, III, and IV. The top and bottom character configurations, or characteristics, are selected by an examination of the character stroke patterns contained therein. The top characteristics can usually be found in quadrants I and II, II and III, or II alone, and the bottom characteristics can usually be found in quadrants III and IV, IV and I, or IV alone. The following is an example of a character separated by quadrants:



To encode the above character, keys P, 7, and 1 (G) are depressed, i.e.



The keys function as follows:

1. The P key represents quadrants I and II.
2. The 7 key represents quadrants III and IV.
3. The sequencing key identifies the row containing the character.
4. The G key, which contains a 1 in the left hand corner, identifies the column containing the character.

The input to the Chicoder is by means of a Friden Flexowriter. When depressed, each of the keys inscribed with stroke patterns generates two unique 6-bit binary codes. The output of the Chicoder is a punched paper tape. Each character is represented by three 6-bit binary codes, or characters. The first 6-bit tape character is a unique code representing the upper stroke pattern, and the second is a unique code representing the lower stroke pattern. The third 6-bit character uniquely identifies the character position in a 5 by 5 matrix. Three bits are used to encode the row, and the other three bits are used to encode the column. A fourth 6-bit tape character indicates character end.

The punched paper tape output of the Chicoder is used as input for language processing from Chinese to English. Each Chinese character uses a Chicode, which contains four characters that are combinations of letters and numbers to make up the lexicographic entries. The first character represents the upper stroke pattern; the second character represents the lower stroke pattern; the third character represents the row; and the fourth character represents the column. A Chinese lexicographic entry of more than one Chicode can be made up, i.e., a Chinese word may be composed of more than one Chinese character, and the Chicode serves as the unique code for each character.

The primary function of the Chicoder in Chinese to English machine translation is to generate from any given Chinese text, a punched paper tape for input to the language processor. Its second function is to provide identifying tags for Chinese characters in the lexicographic entries. This is of prime import in the complete system of language processing, from the input of Chinese to the output of English.

2.2 DICTIONARY ENTRY FORMAT AND ITS RELATIONSHIP TO THE LINGUISTIC PROCESSING SYSTEM

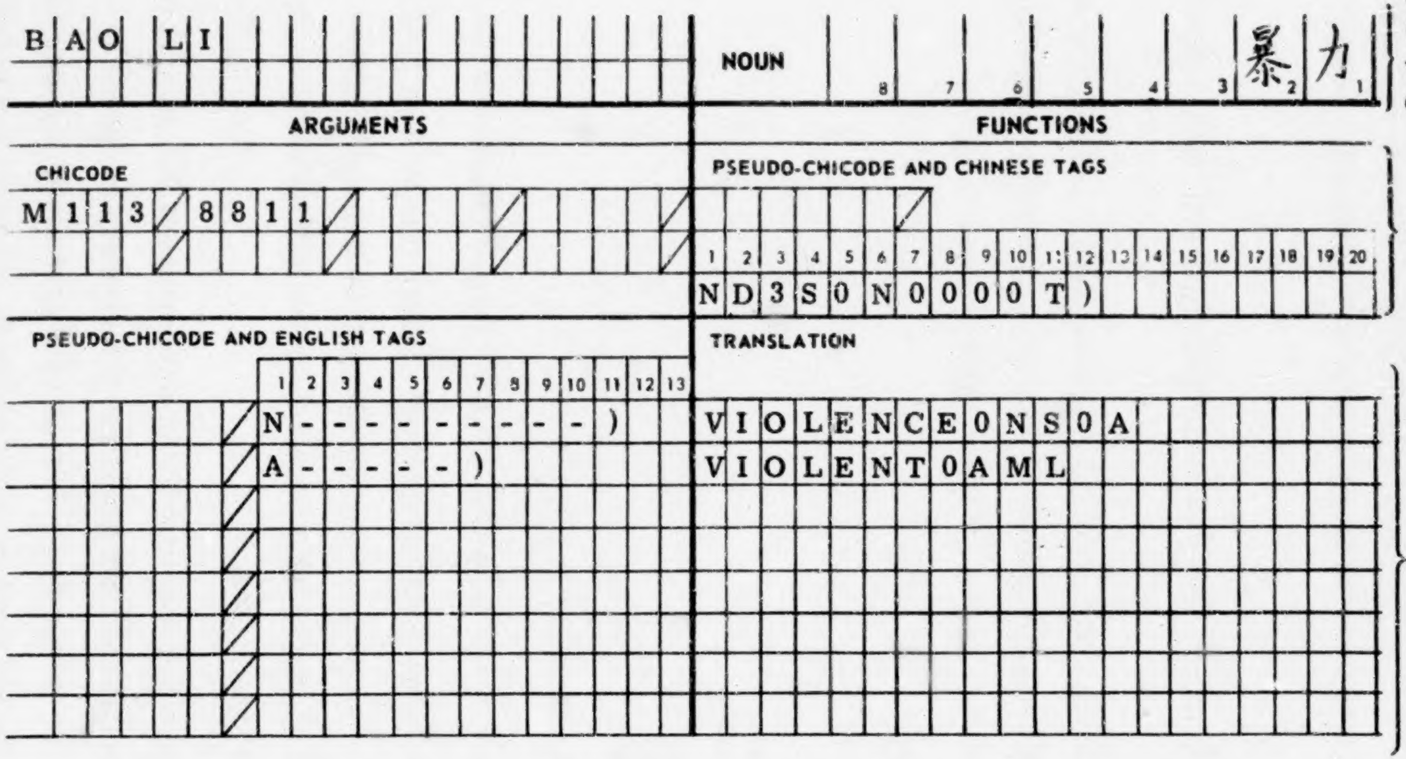
Over 15,000 lexicographic entries consisting of general terms and political science terminology have been completed on a linguistic analysis level, first on card form and then on magnetic tape. The information on each card includes the following items:

1. Romanization (Pinyin)
2. Chinese characters
3. Chicodes
4. Chinese grammatic tags
5. English grammatic tags
6. English translation stem
7. English inflection table tags.

A sample dictionary entry card is shown in Fig. 2-1 and explained in Section 2.5.3.

Selection of lexicographic items from all the word classes (see Section 3) of which the entries are composed is made with the aid of the "Chinese-English Dictionary of Modern Communist Chinese Usage"[5] for the political and general terminology, as well as with the fruits of research and study of general texts listed in Section 8.

The lexicographic entries are divided into three dictionaries (I, III, and IV) for linguistic analysis and programming operations. The argument of dictionary I consists of Chicodes, and the function consists of Chinese grammatic tags. The Chinese grammatic tag may contain ambiguities such as V/N (see Section 3.2.1), the verb and noun tag words for which are listed. Linguistic analysis is accomplished through the use of linguistic rule tables contained in dictionary II, which


B A O L I
NOUN
暴力
ARGUMENTS
FUNCTIONS
CHICODE
M 1 1 3 / 8 8 1 1
PSEUDO-CHICODE AND CHINESE TAGS
N D 3 S 0 N 0 0 0 0 T)
PSEUDO-CHICODE AND ENGLISH TAGS
N - - - - - - - - - -)
A - - - - - -)
TRANSLATION
V I O L E N C E 0 N S 0 A
V I O L E N T 0 A M L

For Reference (romanization letters, word classification, and Chinese characters)

For Dictionary I

For Dictionary III

Fig. 2-1 — Sample dictionary entry card

will be discussed later in this section. The argument of dictionary III consists of Chicodes and English grammatic tags for each word applicable for the entry, and the function consists of English translation stem and English inflection table tags. The English grammatic tags result from altering the original grammatic tags to the correct English grammatic form through the processing of linguistic rules. The English inflection table tags indicate the appropriate English ending for nouns, adjectives, adverbs, and verbs. The noun inflection tables give singular and plural formation and singular and plural possessive formation. Verb inflection tables give tense, infinitive, negative, and auxiliary formation. Adjective and adverb inflection tables give regular, comparative, and superlative formation. For a detailed explanation of the use of English tag tables, see Section 5.

2.3 LINGUISTIC RULE ENTRY FORMAT AND ITS RELATIONSHIP TO THE LINGUISTIC PROCESSING SYSTEM

The linguistic rules are the results of the semantic and syntactical analyses and of the incorporation of the analyses into a machine translation system. These rules are grouped into six major linguistic passes, and each rule is illustrated in symbolic notation on a card whose left and right sides indicate the argument and the function respectively.

Some examples of linguistic rules are:

1. $\mathrm{AJXXXX}_1 + \mathrm{OA}_2$ (argument) → ADXXXX, (function)
2. $\mathrm{ADXKOU}_1 + \mathrm{V} \sim {}^{16}/\mathrm{N} \sim {}_2^{10}$ (argument) → $\mathrm{ADXKOU}_1 + \mathrm{V} \sim {}_2^{16}$ (function)

The explanation of linguistic symbolic notation is as follows:

1. X = anything
2. / = ambiguity division of tag words
3. + = division between one grammatic tag and the other
4. ~ = up to and including the stated number of subtags
5. → = division between argument and function.

The subscripts are word order indicators. A complete explanation of linguistic symbols is presented in Appendix A.

Each major linguistic pass utilizes one or more sets of linguistic rules. These rules are matched against the processing sentence to find a match by word sequence. Special programming operations and lookup techniques are used to alter the original match (the argument) into the function. These sets of rules are dictionary II in the linguistic processing system. They are important in that they furnish the translation system the refinement of language processing. They serve to resolve ambiguities, to group major phrases and attributive structures, to connect the subject with the predicate, and to give inflection.

2.4 LINGUISTIC PROCESSING SYSTEM

In explaining the linguistic processing system, we proceed from the input of Chinese characters to the output of English words. A typist takes a Chinese text and punches from the Chicoder a paper tape, which is used as input for the language processor. A search is made to determine the beginning and end of the sentence, since the linguistic analysis process presently deals with only one sentence at a time. A sentence is indicated by the segmentation indicators KPI, which indicates the beginning of the sentence, and KPT, which indicates the end of the sentence. Dictionary I is utilized and the information from that dictionary for each word of the sentence is extracted and read into the active memory. A processing sentence is illustrated as follows:

- KPI Chicodes/Chinese grammatic tags (first word) Chicodes/Chinese grammatic tags (second word) Chicodes/Chinese grammatic tags (third word) KPT

The processing sentence is isolated, and Chinese grammatic tags are introduced into the linguistic processing operation. Dictionary II, consisting of linguistic rule operations, is then introduced. The sentence is then examined, patterned, and reordered by programming and table lookup techniques. The major programming operations consist of insertion, deletion, reordering, masking, and phrase segmentation. The words on which these operations are performed may be tag words, English words, Chicodes, or translation words.

The linguistic rule processing is composed of six major linguistic passes, which are different from programming passes. The linguistic passes deal with symbolic notation of rules and explanations of operations for these rules, while the purpose of programming passes is to reduce machine time to a minimum by grouping the linguistic rules according to different levels of operation.

Table 2-1 is a list and explanation of the linguistic passes. The major goal of pass 1 is to utilize as much as possible the classes of words that function linguistically as phrase or structure initial or terminal indicators. From these words, major phrases are found and formed into syntactical patterns in subsequent passes. Words with more than one tag word (V/N, A/N, etc.), which we call "ambiguities," are resolved as much as possible, depending on the position of the word relative to other tag words or indicators.

A series of operations is initiated to determine the indicators of verb phrase patterns. In

				Pass 5	Pass 6
Primary phrase segmentation, ambiguity resolving, verb phrase recognition, etc.	Noun phrase recognition (NP)	Attributive structure segmentation and recognition	Relative clause recognition (NP-R)	Scanning of syntactic structure	Choosing of appropriate English words
Subpasses	Subpasses	Subpasses	Subpasses	Subpasses	Subpasses
1A. Primary phrase segmentation—punctuation 1B. Scanning of special structures 1C. Masking of ADRIXX 1D. Possible con* ambiguity resolving 1E. Segmentation by cons 1F. Verb phrase operations (VP): • Indicating verb phrases • Reordering of adverbs, insertion of tense and negatives into verb tags • Forming of verb phrases 1H. Ambiguity resolving II	2A. Searching for and forming simple noun phrases 2B. Reordering and forming complex noun phrases	3A. Collocative structure (LP): • Indicating initial and terminal of collocative structures • Reordering of collocative structures • Forming of collocative structures 3B. Noun phrase containing a collocative phrase (NP-L): • Reordering of noun phrase-collocative • Forming of noun phrase-collocative 3C. Relocation of collocative structure 3D. Connominal phrases (IP): • Indicating terminal of connominal phrases • Reordering of connominal phrases • Forming of connominal phrases • Connominal translation lookup and relocation of connominal phrases	4A. Indicating initial and terminal of NP-R clauses 4B. Reordering of NP-R clause 4C. Forming of NP-R clause	5A. Linking of verb and its object 5B. Linking of subject and its predicate	6A. Grammatical table lookup 6B. English word inflection lookup

* The "cons" are collocatives, connominals, con-adjectival-adverbials, and conjunctions.

pass 1, reordering of adverbs within the verb phrase and tense and modal verb operations are recognized and indicated. The head word of the verb phrase is then singled out with an indicator.

In pass 2, the resulting sequence of tag words is matched against a list of noun phrase patterns. When a match is found, a noun phrase is recognized and the appropriate head word of the phrase is singled out with an indicator. The words within the noun phrase are then rearranged according to English word sequence.

Pass 3 detects attributive structures, which include phrases and subordinate clauses, and finds the major phrase to which these attributes are related. The attributive structures are then isolated so that a syntactical pattern becomes apparent.

Pass 4 consists of a series of operations to discover the existence of noun phrase-relative clauses in the sentence. When a noun phrase-relative clause is found, the word order is rearranged according to equivalent English word sequence. The head word of this structure is again singled out and treated as a nominal in the syntactical structure.

Pass 5 links the head words of all major phrases in the sentence and determines the sentence pattern according to linguistic tags of each head word. This pass completes operations on all the tables in dictionary II.

The final pass, pass 6, selects the grammatically and semantically appropriate English word from a group of translations for each Chinese word. This pass utilizes dictionaries III and IV.

The details of each linguistic pass are further amplified in the following paragraphs.

2.4.1 Pass 1

Pass 1 is divided into eight subpasses. Pass 1A sections the sentence according to the existing punctuation marks. This is utilized to indicate the possibility of initial and terminal points of phrases or clauses. When a punctuation mark is found, two indicators are usually inserted before and after the punctuation mark. The primary phrase segmentation table has two types of punctuation indicators. The KC indicator is inserted before and after punctuation marks within a sentence, such as comma and semicolon. The KP indicator is inserted before and after punctuation marks, such as question mark, period, paragraph indicator, and exclamation mark, that indicate the end of a sentence. For example:

1. PC → KCI + PC + KCT
2. PP → KPT + PP + KPI

Pass 1B picks out special structures, such as comparative structures or interrogative sentences, without going through subsequent passes, since the syntactical patterns for these structures need further analysis.

Discontinuous structures that are used for subtle shades of meaning for words having no appropriate English equivalents are scanned and masked. For example:

- 黑夜是有尽头的 (dark night has ending).

The more appropriate translation would be: "The dark night does have an ending." For the present 是 is masked, since it is treated as a verb, and emphasis is not analyzed in this phase of study.

Another example is:

- 有勇气的青年 (brave youths).

It is more appropriate to use the adjective form of the noun for 有 ... 的 with certain nouns between than to consider this as a noun phrase-relative clause. In similar phrases, such as 政治上的问题 (political problems), it is better to use the adjective form of 政治 (political).

Pass 1C picks out adverb independent words that do not affect the rest of the sentence, and they are masked for subsequent scanning (see Section 3.3).

In Pass 1D, the functions of specific connominal, collocative, and converbal class words are determined by examining their immediate environment, and they are indicated accordingly. These words can therefore be utilized as phrase or structure initial and terminal indicators. For example, if 要 is immediately followed by a collocative initial, it is used as a converbal and not as a verb. Those words whose functions cannot be decided on until the scanning of a greater structure will be dealt with later (see Sections 3.13 and 3.15).

Pass 1E searches for the initial or terminal indicators for prepositional phrases or attributive structures. When a phrase or structure indicator is found, the KI and KL indicators are inserted before and after phrase or structure tag words such as LIX and IXXXX. A minor indicator (KM) is inserted after the phrase indicator word if it is an initial indicator or before if it is a terminal indicator. Initials and terminals must be equal in number, and they are paired by linguistic routines.

Pass 1F deals with the resolving of ambiguities of words with more than one tag word, a circumstance that depends on the immediate environment of the word. The rules may be reapplied in the sentence for these ambiguities. For example:

1. $AAXXXX_1$ (adjective adverb) + OA_2 (地) + V_3 = $ADXXXX_1$ (adverb) + V_3
2. $AJXXXX_1$ + V/N_2 = $AJXXXX_1$ + N_2

Pass 1G searches for and isolates main verbs. A series of operations, including verb phrase segmentation, adverb reordering, and inserting tense into verb grammatic tags, is made. The verb is then singled out for subsequent passes (see Section 4.2).

Pass 1H resolves words with more than one tag word by examining their environment. This operation may be performed after each major phrase operation is completed.

2.4.2 Pass 2

Pass 2 is concerned with the recognition of simple and complex noun phrases and with the isolation of the head word from its attributes in the context. The head word is utilized for subsequent passes for prepositional phrases, etc. In this pass, the number of the head word is affected by its attributes, such as arithmates or numerals, that have number tags. Words are reordered according to equivalent English word sequences. The detailed operation for this pass is described in Section 4.1.

2.4.3 Pass 3

This pass searches for initial or terminal points of collocative structures and connominal phrases. When the initial or terminal point is found for the structure, indicators are inserted to separate the structure from the processing sentence. The elements within the structure are reordered for the proper translation. After this process, the tag words within the structure are masked except for the connominal tag word and the initial and terminal structural indicators. The connominal tag word is then referenced to the related verb for the proper connominal translation. The collocative structures and connominal phrases are then syntactically reordered to follow the related verb or verb and object if required. Each of these steps is described in detail in Sections 4.3 and 4.4.

2.4.4 Pass 4

In this pass, the environment of the noun phrase-relative clause is examined to determine the initial and terminal points of the noun phrase-relative clause. When these are found, initial and terminal indicators are inserted before and after the clause to segment the structure from the sentence for subsequent operations. To derive the proper English translation of the noun phrase-relative clause, tag words within it are reordered and English words are added where necessary.

All words are then masked except the head word, which is to be singled out for the verb linking pass (see Section 4.5).

2.4.5 Pass 5

By this time, all attributive structures have been masked. The modifiers of noun phrases (including noun phrase-relative clauses) and of verb phrases have been masked, and only the head words of these phrases remain in the processing sentence. At this stage, all attributive structure initial and terminal indicators as well as verb phrase initial and terminal indicators (KXX) are masked, with the exception of KCI, KCT, KPI, and KPT. Therefore, only major punctuation initial and terminal indicators and head words of major phrases remain in the processing sentence. The operations are then initiated to link the noun head word with the verb head word to give the verb person and number. To derive certain operations, a series of rules indicating the lookup of the next head word is necessary for this phase (see Section 4.6). After this process is completed, the processing sentence is ready for the selection of the proper translation according to the English grammatic tag of individual Chinese words (Chicodes).

2.4.6 Pass 6

In the first five passes, the Chinese grammatic tags for each word are altered to the appropriate English grammatic tags. In pass 6, dictionary III is utilized in matching the appropriate English grammatic tags to find the correct stem of the English translation. The correct English table tags are used in looking up the ending for the correct form. A set of special operations is initiated to determine the proper form for the inflection specified in the English grammatic tag. Dictionary IV gives the tables for these appropriate forms so that programming operations can be specified to generate operations that search for the correct endings (see Section 5). Auxiliary words and/or endings are then attached to the word stem to generate the English equivalent for the Chinese word.

2.5 COMPUTER EXPERIMENTATION

The computer used in the present system for Chinese to English machine translation (MT) is a Digital Equipment Corporation PDP-1 with auxiliary disc storage. The system utilizes programs that allow the application of automatic lookup, content addressed tree structure, and context associative type techniques. Several special operations have been programmed to facilitate a simulated content addressed table lookup method that is associative but allows nesting and structuring for phrasing, etc. These initial operations are experimental in nature and are being used both for proving the approach and for ultimate system design specifications. Chinese inputs and outputs are encoded on the Chicoder and a paper tape is produced. This tape is processed by the experimental MT system; outputs are punched on paper tape with subsequent English printout on a Friden Flexowriter.

Three dictionaries are prepared according to the longest match principle to allow a simulation of content addressing. Dictionary I accomplishes word segmentation and converts Chicodes of Chinese words into pseudo-Chicodes (semantic tags) with Chinese (grammatic) tags. Dictionary II converts Chinese tags into English grammatic tags through operation of different linguistic rules. Dictionary III leads semantic tags and English grammatic tags to a final translation in English. A set of subdictionaries, i.e., dictionaries for English inflection table tags, will later enable automatic derivation of the proper auxiliary forms and word endings.

2.5.1 Processing Procedure

The processing procedure contains three passes using three dictionaries respectively.

Pass 1 — Chicodes to Semantic and Grammatic Tags. Dictionary I tape is written onto the disc by an automatic loading program. The tape of Chinese text (in Chicodes) is then read in to be processed by the content addressed lookup technique, i.e., when the Chicodes (argument) of a Chinese word are associated with an entry in dictionary I, they will be replaced by the corresponding semantic tags and Chinese grammatic tags (function). The output of pass 1 is thus a string of semantic and Chinese grammatic tags. (See Appendix B for listings of Chinese text, dictionaries, and outputs of the experiment.)

Pass 2 — Applications of Linguistic Rules. After dictionary II is loaded onto the disc, the output of pass 1 is read in to be processed. Arguments of linguistic rule entries are associated with the input stream, which is then modified by operation codes indicated in the functions. An output tape of this pass is then produced as the input of the next pass. The output of pass 2 is a string of semantic tags and English grammatic tags that are to be associated with the arguments in dictionary III. Special operations are used here in facilitating associative techniques to accomplish the linguistic operations. The operations are described and explained in Section 2.5.4.

Pass 3 — Semantic Tags and English Grammatic Tags to English Translations. After dictionary III is loaded onto the disc, the output of pass 2 is read in to get the final translation. This pass uses only the content addressed lookup technique (with chained stems and endings). The output of pass 3 is the English translation of the original Chinese text. However, the English translation output for nominals, adjectival adverbials, and verbals appears in the stem form of the vocabulary (see Section 5) with subsequent attachment and insertion of proper endings and auxiliary forms through the use of subdictionaries.

2.5.2 Dictionary Entry

Dictionary Format. An entry card has been designed in special formats for dictionaries I and III. This card is arranged for the convenience of direct punching without further coding work (see Fig. 2-1). The entry card is divided into three portions. On the top portion of the card, blocks on the right are for Chinese characters while blocks on the left are for their romanization (Pinyin) letters. The space between Chinese characters and their romanized letters is left for indicating word classification, i.e., noun, verb, etc. In the middle portion (for dictionary I), Chicodes of Chinese characters (argument) are of the left, and pseudo-Chicode (semantic) and Chinese (grammatic) tags (function) are on the right. In the lower portion (for dictionary III), semantic and English grammatic tags (argument) are on the left while English translations and inflection table tags (function) are on the right.

Chinese Characters. Chinese characters of a word are copied in blocks. They are picked up mainly from the "Chinese-English Dictionary of Modern Communist Chinese Usage."[5] Other sources include articles from "Hong-Qi" magazine,[14] the "People's Daily,"[23] and other publications.

Romanization Letters. These are the transliteration of Chinese characters. The romanization procedure used is the Pinyin system of Communist China rather than the Wade-Giles system. The "Chinese-English Dictionary of Modern Communist Chinese Usage"[5] is used as a standard reference for romanization Pinyin.

Chicodes. The Chicode of each Chinese character can be found from Chicoder typing or from the reference manual.[22] Each Chicode contains four characters that are combinations of letters and numbers. A slash is used as the delimiter of each Chicode, e.g., ck12/t711/.

Pseudo-Chicode (Semantic Tag). For efficient processing, the pseudo-Chicode or semantic tag is designed to represent variable length Chicodes of each Chinese word. Each sematic tag has a fixed length of six characters that are combinations of letters and numbers with a slash as the delimiter, e.g., vaab99/. The semantic tags are very important in the application of linguistic rules during the computer processing, which will be described in Section 2.5.5.

Chinese Grammatic Tags. Chinese tags of each word have variable lengths due to different word classifications, e.g., 10 tags for a noun and 16 tags for a verb. These tags are symbols of grammatic analyses of a Chinese word.

English Grammatic Tags. English tags of each word also have variable lengths similar to Chinese tags. These tags are English translation grammatic indicators of, e.g., word classification, number, and person. There may be more than one classification (part of speech) for one Chinese word. There may therefore be more than one set of English tags to indicate different word classifications.

Translation. This refers to the English translation of one Chinese word. One Chinese word may be used for different word classifications (with different sets of English tags) and one Chinese word may therefore have more than one translation.

2.5.3 Linguistic Rules

Linguistic Passes. According to the procedure of linguistic analyses, linguistic rules are classified into six linguistic passes with some subpasses. (These are not to be confused with the dictionary passes, since much of the computer processing is internal in nature.) They are listed and described elsewhere in this section as well as in other sections of this report. Linguistic rules are referred to as dictionary II (pass 2 during processing) in the explanation of this experiment. All linguistic passes except linguistic subpass 6B are included in processing pass 2. The linguistic subpass 6B, "English word inflection lookup," is left at this stage. Further effort will be made to continue full implementation of table tags for choosing the proper ending of English words.

Writing Format of Linguistic Rules. Linguistic rules are written in a format similar to that of the chemistry equation, with arguments on the left and functions on the right. Examples of some linguistic rules are shown below. All letters and numbers in the equations are either Chinese tags of a word (e.g., aa-----) or linguistic tags for analyses (e.g., kct). Semantic tags of each word are not shown in the equation. Subscripts to sets of tags indicate their sequence. Superscripts found at the end of some sets of tags indicate the total number of tags in the set.

1. $\text{AAXXXX}_1 + \text{HM}_2 \rightarrow \text{AJXXXX}_1 + \text{HM}_2$

This rule is intended to solve a simple adjectival-adverbial ambiguity. The equation means that when there is an adjectival-adverbial ambiguity (AAXXXX) that is immediately followed by a special class word (HM), the adjective form (AJXXXX) is used.

2. $\begin{bmatrix}\text{KCT}\\ \text{KPI}\end{bmatrix}_1 + \$\overline{\text{VTIC} \sim {}_2^{16}} + \text{UP TO NEXT}_3 + \text{AJXXXX}_4 + \text{HM}_5 + \text{UX}_6 + \text{N} \sim {}_7^{10} \rightarrow \begin{bmatrix}\text{KCT}\\ \text{KPI}\end{bmatrix}_1$

$+ \$\overline{\text{VTIC} \sim {}_2^{16}} + \text{UP TO NEXT}_3 + \text{UX}_6 + \text{AJXXXX}_4 + \text{N} \sim {}_7^{10}$

This example shows the matching of alternative case and discontinued string, the reordering of word sequence, and the deleting of a word. The bracketed expression $\begin{bmatrix}\text{KCT}\\\text{KPI}\end{bmatrix}$ means that either KCT or KPI will appear at that position. The dollar sign with overscore, $\$\overline{}$, means that anything else except what is under the overscore will be applied to this rule. UP TO NEXT represents the discontinued portion of the string, and means the connection of two halves of the rule. The whole equation means that when either KCT or KPI is immediately followed by anything else except VTIC ~ ^{16}, and followed by some other words that are in turn followed by a string of AJXXXX, HM, UX, and N ~ ^{10}, the word AJXXXX should be moved to the position between UX and N ~ ^{10} and the word HM should be eliminated.

$$3.\quad \text{US}_1 + \text{UP}_2 + \text{UP}_3 + \text{UP}_4 + \text{NT3S0N00U0}_5 \rightarrow \text{NT3S0N30U0}_5 + \underline{\text{of} + \text{UA}_1 + \text{UA}_2 + \text{UA}_3 + \text{UA}_4}$$

This is an example of noun phrase reordering. When there are four numerals (US or UP) that are followed by a noun of time (NT3S0N00U0), they should be presented in arithmetic form (UA) and moved to the position after the noun with "of" in front of the numerals. After reordering, the underlined portion in the function (of + UA + UA + UA + UA) is masked until final translation of English. The nominal tag 7 is therefore changed from 0 to 3 to indicate the formation of a nominal of time phrase.

2.5.4 Operation Codes for Contextual Associative Method

Operation codes indicated in dictionary II will perform the special matching operation and the linguistic changes required by translation. These operations will allow the input data to be manipulated in different ways to obtain the desired form. Ambiguities in word classification will be resolved. Noun phrases, verb phrases, and adverbs will be reordered and put into proper position. The subject, the main verb, and the object of a sentence will be detected and connected to each other. These operation codes are described and explained individually as follows. However, examples of their uses will be given later.

Don't Care Code (-). A hyphen in the argument will associate with any character in the input string at that position. One hyphen will match only one character. The hyphens in the argument are therefore used as unspecified characters to associate with corresponding characters in the input. (The don't care code is identical with X in linguistic analysis.)

Up To Next (→). The right arrow will appear in the argument for connecting two separate parts of a rule. This operation is created because some linguistic rules require the association of two unrelated strings to provide sufficient information. The gap in between may contain variable length data, and it is therefore not known how many don't care codes are needed. This up to next code will fill the gap and connect the two separated parts to enable a complete association.

Save Input (.). The period is used in the function to save the associated portion of the input as indicated in the corresponding argument. This code will usually appear at the beginning of the function to alter the input string. Without this code at the beginning of the function, the associated portion of input will be deleted from the input string.

Rho Stuffing (‾). The overscore in the function will insert characters underneath and followings into the input string. Characters to be inserted start from the one under the overscore and end at the one before any operation code or blank.

Shift Code (• or •m). The middle dot indicates that the assumed pointer in the input string is to be shifted to the right. A letter m immediately following the middle dot will change the shifting direction to the left of the pointer. Numbers following the shift code indicate how many positions to be shifted. The assumed pointer is the position in the input string where the next table lookup will start. The pointer will appear in the intermediate printout as an underscore (_).

Ignore Codes (⊃, ∧, <, >, [,]). There are six sets of ignore codes designed to ignore portions of the input for future table lookup. These ignore codes are used only in pairs of the same code, e.g., [and [, < and <. The first ignore code of the pair indicates the initial of the ignored portion while the second ignore code indicates the terminal. One pair of ignore codes could be included in another pair of ignore codes if necessary. (The ignore operation is identical with the masking function in the linguistic analysis.)

Masking Code (,). The comma is used in the function to mask or skip the unchanged character in the argument, i.e., the comma in the function will save the corresponding character in the argument as it is, therefore retaining it in the input in original form. One comma will skip only one character.

Save Copy (↑). The up arrow in the function will save part of the argument and cause the saved portion to be moved to the desired position. Masking codes usually follow the up arrow to indicate the corresponding characters that will be moved.

Restore Copy (+). A plus sign following a shift operation in the function will restore the part saved to the designated position. This operation code is used in conjunction with a save copy code and a shift code to accomplish the reordering of data strings.

Reset Pointer (~). A tilde at the end of function will reset the pointer at the beginning of a sentence. This code is often used at the completion of rule application to stop further searching of the rest of the sentence, and to restart the searching from the very beginning of the sentence.

2.5.5 Coding of Linguistic Rules

Linguistic rules are coded in dictionary II in a pseudo language for processing. The three examples cited in Section 2.5.3 are brought over to show the actual coding of individual rules. The coding of example 1 is:

```
              Argument              Function

          aa----)------/hm)          .x̄1=

          x1=aa                      aj ~
```

These two dictionary entries will accomplish the simple ambiguity rule in example 1. The hyphen stands for anything, such as a letter, a number, or a symbol. When the argument of the first entry is addressed, its function specifies the saving of the associated portion of the input stream and the moving of the pointer to the left of the argument, indicated by a period, and stuff in x1= at the location of pointer, indicated by an overscore. The string in the I/O buffer will then show as follows, with the underscore designated as the pointer:

- x̲1=aa----)------/hm)

The second entry will associate with the above string and replace x1=aa with aj. The pointer will be moved back to the very beginning of the processing sentence (indicated by ~). The string will show as:

- aj----)------/hm)

The coding of example 2 is:

```
              Argument                                  Function

   kct (------/vtic------------)→-------/          .•999999998
   aj----)------/hm)------/u-)------/
   n---------)

   kpi(------/vtic------------)→-------/           .•999999998
   aj----)------/hm)------/u-)------/
   n---------)

   kct(→-------/aj----)------/hm)------/           .•4↑,,,,,,,,,,,,,•992+•m992x̄20=
   u-)------/n---------)

   kpi(→------/aj----)------/hm)------/            .•4↑,,,,,,,,,,,,,•992+•m992x̄20=
   u-)------=n---------)

   x20=------/hm)
```

Example 2 needs five dictionary entries to accomplish the rather complicated requirement. In the first entry, any portion of the input stream meeting these conditions will associate with the example 2 rule, which began with KCT + VTIC ~ ^{16}. Its function specifies the saving of the associated input stream (indicated by a period) and the shifting of the pointer to the right (indicated by a middle dot) by 80 characters (indicated by 999999998). This will meet the requirement of example 2 that if KCT or KPI is immediately followed by VTIC ~ ^{16} and the rest, this rule will not apply. The second entry is the same as the first entry except that KPI replaces KCT to ensure that either case will not be applied.

Since exceptions of example 2 have been taken care of, the third and fourth entries will include all cases applicable to this rule. The up to next code (→) will connect two separated parts of the rule. Their functions indicate the saving of that portion of the input stream and the shifting of the pointer to the right by 4 characters (indicated by • 4) and the moving of the following 14 characters (indicated by †,,,,,,,,,,,,,,) to the right by 20 positions (indicated by • 992) and the restoring of them (indicated by +). The pointer is then shifted left by 20 positions (indicated by • m992) and x20= is inserted. After operations of this entry, the original string (argument of third or fourth entry) will appear as:

- kct(→x̲20=------/hm)------/u-)------/aj----)------/n---------)

or as:

- kpi(→x̲20=------/hm)------/u-)------/aj----)------/n---------)

The pointer is now at x̲ and the argument in the fifth entry will therefore be addressed. Because there is no save code (.) at the beginning of its function, the argument will be eliminated from the data string. The tilde in the function will reset the pointer to the beginning of this sentence for other table lookups. The string after the application of the fifth entry will be shown as:

- kct(→-------/u-)------/aj----)------/n---------)

or as:

- kpi(→-------/u-)------/aj----)------/n---------)

The coding of example 3 is:

Argument	Function
------/us)------/up)------/up)	.,,,,,,/ua),,,,,,/ua),,,,,,/ua)
------/up)------/nt3s0n00u0)	,,,,,,/ua)•994x̄22=

```
x22=00u0)                          30u0)zofzzz/z)•m999x̄23=

x23=------/nt3s0n30u0)zofzzz/z)    .x23=↑,,,,,,,,,,,,,,,,,,,,,,,,,,•m99998+•99
                                   <•999994x̄24=

x24=x23=                           <~
```

This rule of reordering noun of time phrases is facilitated by four entries in dictionary II. When the argument of the first entry is matched by the string in the I/O buffer, its function indicates the saving of the argument (indicated by a period) with masked portion to be protected (indicated by commas) and unmasked portion to be replaced. The pointer is then shifted to the right by nine and four characters and x22= is inserted there. After the application of the first entry, the string will show as:

- ------/ua)------/ua)------/ua)------/ua)------/nt3s0nx̲22=00u0)

The last portion of the string will match the argument of the second entry and is to be replaced by 30u0)zofzzz/z). The pointer is then moved to the left (indicated by • m) by 9+9+9 characters to insert x23=. The string will show as:

- ------/ua)------/ua)------/ua)------/ua)x̲23=------/nt3s0n30u0)zofzzz/z)

The portion started from x23= will associate with the argument of the third entry. The function of the third entry instructs the saving of the associated portion of the input stream (indicated by .) and the moving of the 27 characters (indicated by↑,,,,,,,,,,,,,,,,,,,,,,,,,,) that follow x23= to the left (indicated by • m) by 9+9+9+9+8 characters. It then instructs to shift the pointer to the right (indicated by •) by 9+9 characters to insert the ignore code (<) and again to shift to the right by 9+9+9+9+9+4 characters to insert x24=. The string will become:

- ------/nt3s0n30u0)<zofzzz/z)------/ua)------/ua)------/ua)------/ua)x̲24=x23=

The last portion of the above string will associate with the argument of the fourth entry. Its function indicates the replacing of the associated input stream segment by another ignore code (<) and the moving back of the pointer to the very beginning of the sentence (indicated by ~). The final appearance of the string will be:

- ------/nt3s0n30u0)<zofzzz/z)------/ua)------/ua)------/ua)------/ua)<

2.5.6 Conclusion

To give a precise picture of computer processing results, a two-sentence paragraph of

Chinese text is chosen as an example for illustration. The Chinese text is prepared and processed according to the procedures described above. The Chinese text in both Chinese and Chicodes, three dictionaries, and outputs of each pass are listed in Appendix B. However, for easier understanding, dictionary II (linguistic rules for the sample paragraph) is listed in linguistic writing format rather than in the computer pseudocoding language.

Briefly, this experimentation has successfully adapted the linguistic analyses to computer implementation of language processing. This developed and tested machine translation system is not a simple word by word machine translation but will utilize and apply accumulated efforts of the linguistic research to give an output closer to that of human translation. The quality of the machine translation can always be improved by enlarging and revising morphological and syntactical dictionary entries, which are the direct result of the linguistic research.

The software system employed is open ended in every sense, in that rules can be added for exceptional cases and rules can be added to apply to greater context without requiring special programming and without causing any conflict (unless it supersedes a rule).

Although this system is designed for Chinese to English translation, its principle and methodology could be effectively applied to the inverse translation, i.e., English to Chinese, or to other language translations, e.g., Russian to English or German to English.

3. MORPHOLOGICAL ANALYSIS

An intensive study of the morphology of the modern Chinese language was made during the past year. In the course of this study, many factors in the Chinese language were considered. Monosyllabic as well as polysyllabic Chinese words were examined for their functions in relation with other words. Both semantic and word class ambiguities were taken into account. Inflections for the three primary morphological classes (nominals, verbals, and adjectival adverbials) were considered in relation to monosyllabic and polysyllabic words whose definitions are enhanced by these inflections. The relative importance of punctuation marks in the segmentation of phrases and clauses was examined. In the composition of morphological classes, ample room was allowed for adjustments, modifications, and additions.

Each major morphological class has a series of subtags, and each subtag gives specific grammatic information for a particular word in that class. For example, the adjectival-adverbial class has subtags that denote degree, tense, and type of word modified by adverbials. Up to this point, 14 morphological classes have been defined, and room for addition of subtags to each class and for formulation of more word classes has been allowed. Tag 1 is the designation for each major morphological class, the subtags for which are illustrated in the general table (Table 3-1) and explained in this section of the report. The specific tables for the Chinese morphological tags are shown in Tables 3-2 through 3-8.

3.1 MORPHOLOGY TAGS AND EXPLANATIONS

The 14 word classes outlined in the general and specific tables for Chinese morphological tags are utilized in making lexicographic entries. Specifically, the major word class and its subtags are listed as Chinese grammatic tags for each entry in the dictionary format. Each Chinese word may, of course, have one or more word classes for morphological analysis. Table 3-9 gives the symbol, the terminology as used in our linguistic analysis, and the general definition for each major word class (tag 1).

Table 3-1 — General Table of Chinese Morphological Tags

Tag 1	Tag 2	Tag 3	Tag 4	Tag 5	Tag 6	Tag 7	Tag 8	Tag 9	Tag 10	Tag 11
Nominals (N)	Type	Person	Number	Capitalization	Gender	YD&S	Address	Definity	Translatability	
Adjectival adverbials (A)	Type	Degree	Quality of adverb	Tense	Translatability					
Verbals (V)	Type	Object	Complement after object	Pretransitive (I)	Preverbal 对 (IC)	Preverbal 向(ID)	Postverbal 为 (IK)	(Blank)	Postverbal, other I's	Preverbals 同, 和, 与, 跟 (IA)
Arithmates (R)	Number									
Synonomes (S)	Translatability									
Numerals (U)	Number									
Conjunctions (J)	Type									
Auxiliaries (G)	Type									
Con-adjectival-adverbials (O)	Type									
Punctuation (P)	Type									
Special (H)	Type									
Connominals (I)	Type	Regularity	Position	Quality						
Converbals (E)	Type	Subtag	Tense							
Collocatives (L)	Type	Translation classification								

Table 3-2 — Specific Table of the Nominal Class Tags

Tag 1—Kind	Tag 2—Type	Tag 3—Person	Tag 4—Number	Tag 5—Capitalization	Tag 6—Gender	Tag 7—YD&S
N = nominal	M = pronominal	1 = first	S = singular	0 = none	N = neuter	0 = none
	C = country	2 = second	P = plural	P = proper	M = masculine	Y = 有 . . . 的 (Y . . . D)
	D = idea	3 = third		M = romanization	F = feminine	
	O = organization					S = 上
	H = human					K = S and Y
	E = human collective					
	I = inanimate					
	B = beast					
	P = place name					
	T = time					
	K = collocative terminal					

8—Address

= none

= address

= family
me

Tag 9—Definity

0 = none

D = definite

I = indefinite

U = numeral

R = relative

N = before numeral

O = ㄹ

Tag 10—Translatability

0 = translatable

U = untranslatable

Table 3-3 — Specific Table of the Adjectival-Adverbial Class Tags

Tag 1—Kind	Tag 2—Type	Tag 3—Degree	Tag 4—Quality of Adverb	Tag 5—Tense	Tag 6—Translatability
A = adjectival adverbial	J = adjective	0 = none	0 = none	0 = none	T = translatable
	D = adverb	R = regular	V = verb	P = present	U = untranslatable
	A = J and D	C = comparative	J = adjective	A = past	
	P = J postal	S = superlative	D = adverb	F = future	
	B = P or D		A = J and D		
			R = V and J and D		
			B = V and D		
			E = V and J		
			I = independent		
			K = I and V		
	S = special adjective	0 = none	A = 与.同,和,跟	1 = 一样	T = translatable
		R = regular	S = 像,如	2 = 相同	U = untranslatable
		C = comparative	V = 离	3 = 远	
		S = superlative	Z = A and S	4 = 近	
				5 = 一起,一同, 一道,一伙	

Table 3-4 — Specific Table of the Verbal Class Tags

Tag 1—Kind	Tag 2—Type	Tag 3—Object	Tag 4—Complement After Object	Tag 5—Pretransitive (I)	Tag 6—Preverbal对(IC)	Tag 7—Preverbal向(ID
V = verbal	I = intransitive	0 = none	0 = none	0 = none	0 = none 1 = to 2 = concerning 3 = in 4 = with 5 = against 6 = from 7 = at P = pretransitive	0 = none 1 = to 2 = from 3 = against 4 = toward 5 = at 6 = with P = pretransitive
	T = transitive B = I and T	A = noun without complement	0 = none	0 = none A = 对 B = 把 C = A and B		
		B = noun with or without complement C = noun with complement	I = infinitive A = apposition W = nominal B = I and A C = I and W D = A and W E = I, A, and W			
		D = noun or infinitive E = infinitive F = noun or clause G = clause	0 = none			
	S = special	1 = 是 2 = 有	0 = none	0 = none	0 = none	0 = none
	W = verbs of adverbial quality	3 = participle 4 = infinitive				

ag 8—Postverbal 为(IK)

0 = none
1 = into
2 = as

Tag 9 –Blank

0 = none

Tag 10—Postverbal, other I's

0 = none
N = into (成)
O = as (作, 作为, 做)
P = N and O

Tag 11—Preverbal 同, 和, 与, 跟, (IA)

0 = none
1 = with
2 = as
3 = from

Table 3-5 — Specific Tables for the Arithmate, Synonome, Numeral, Conjunction, Auxiliary, Con-Adjectival-Adverbial, Punctuation, and Special Class Tags

Arithmate Tags

Tag 1 — Kind	Tag 2 — Number
R = arithmate	S = singular
	P = plural

Synonome Tags

Tag 1 — Kind	Tag 2 — Translatability
S = synonome	T = translatable
	U = untranslatable

Numeral Tags

Tag 1 — Kind	Tag 2 — Number
U = numeral	S = singular
	P = plural

Conjunction Tags

Tag 1 — Kind	Tag 2 — Type
J = conjunction	C = connector

Auxiliary Tags

Tag 1 — Kind	Tag 2 — Type
G = auxiliary	B = must, must not
	Y = should, should not
	N = can
	K = cannot but
	Q = must never

Con-Adjectival-Adverbial Tags

Tag 1 — Kind	Tag 2 — Type
O = con-adjectivial-adverbial	A = 地
	B = 得
	C = 反
	D = 次, 回, 倍, 趟
	E = 成
	F = 百分之
	G = 分之
	H = 第

Punctuation Tags

Tag 1 — Kind	Tag 2 — Type
P = punctuation	C = comma
	P = period
	H = dash
	L = colon
	M = semicolon
	I = paragraph initial
	E = exclamation
	Q = question
	S = series
	D = name division
	B = book title
	T = quotation
	R = parenthesis
	A = deleted passage

Special Tags

Tag 1 — Kind	Tag 2 — Type
H = special word	M = 的
	N = 们
	Z = 之
	S = 性
	H = 者
	D = 等
	U = 多

Table 3-6 — Specific Table for the Connominal Class Tags

Tag 1 — Kind	Tag 2 — Type	Tag 3 — Regularity	Tag 4 — Position	Tag 5 — Quality
I = connominal	A = 同，跟，和，与	R = regular	R = preverb	N = followed by noun or noun phrase
	B = 替，代	I = irregular	P = postverb	V = N, or followed by verb (present participle) and noun
	C = 对	A = with special adjective	B = R and P	S = clause or noun phrase
	D = 向	C = comparison connominal	I = independent of verb	
	E = 給，被		A = preadjective or comparison	
	F = 把，将			
	G = 对于			
	H = 根据，就，按照，按，照，			
	I = 由			
	J = 由于，不管			
	K = 为			
	L = 为了			
	N = 成			
	O = 作，作为，做			
	Q = 特别是，尤其是			
	R = 比，比較			
	S = 像，如，如同，正如，例如			
	T = 从			
	U = 没有，不及			
	V = 离开，离			
	W = 經过，通过			

Table 3-7 — Specific Table for the Converbal Class Tags

Tag 1 — Kind	Tag 2 — Type	Tag 3 — Subtag	Tag 4 — Tense
E = converbal	N = negative P = present participle T = tense indicator	I = initial T = terminal B = I and T	F = future P = present R = progressive E = perfect
	A = passive voice indicator (被) R = relative clause A (所)	I = initial T = terminal	0 = none
	I = important verb indicator	A = 而 B = 来 C = 去	0 = none
	C = complement	3 = untranslatable 6 = translate as it is	0 = none

Table 3-8 — Specific Table for the Collocative Class Tags

Tag 1 — Kind	Tag 2 — Type	Tag 3 — Translation Classification
L = collocative	I = initial	A = 除, 当, 除了, etc. B = 向, 朝, 朝着, etc. C = 从, 由, 自, etc. D = 在, etc.
	T = terminal	A = 北面, 边上, etc. B = 身上, 头頂, etc. C = 当中, 底下, etc. D = 后, 看来, etc. E = 来, 里, 起, etc. F = 說来, etc.

Table 3-9 — Symbols, Terminology, and General Definition for Major Chinese Word Classes

Symbol	Terminology	General Definition
N	Nominals	Nouns (太陽)
A	Adjectival adverbials	Major attributes of nouns and verbs (真实,匆忙)
V	Verbals	Verbs (挽留)
L	Collocatives	Prepositions of discontinuous structure (在. . .上,从. . .以后)
R	Arithmates	Demonstratives (这,那,这些,那些)
S	Synonomes	Classifiers and measurements (个,张,种,尺,寸)
U	Numerals	Numbers (一,二,三,四,百万)
J	Conjunctions	Equal conjunctions (和,以及,或者)
E	Converbals	Words contributing to verbs in regard to tense, aspect, voice, negation (已經,不,被,正在,了,着)
G	Auxiliaries	Auxiliary verbs (应該,千万不可,千万要,必须)
O	Con-adjectival-adverbials	Words that formulate adjectives or adverbs
P	Punctuation	Punctuation marks (. ; .)
I	Connominals	Prepositions (像,为了,对,把)
H	Special words	Special words (的,们,等)

3.1.1 General Tag Terminology

The following paragraphs briefly explain the general tag terminology.

Type. The types of a word include subclasses that denote either semantic or grammatic qualities of differentiation. For example, nominals are differentiated according to meaning; they fall into the classes of human, country, animal, abstract, concrete, place name, etc. Verbs are classified according to grammatic function, such as transitivity with nominal object, transitivity with both direct and indirect object, and intransitivity.

Position. The positions of a word include subclasses that denote its possible relative positions between words. For example, collocative initial precedes collocative terminal (在...中), converbal initial precedes the verb, and converbal terminal follows the verb (将, 了).

Number. Number classification is tagged in accordance with the English translation as to singular and plural, since number inflection is not shown in Chinese. Word classes such as nominal, arithmate, and numeral have number tags.

Translatability. Synonomes have translatability and untranslatability. For example, 个 is untranslatable, while 种 (kind) is translatable. Some adverbs, such as 才 and 还, which are presently difficult to translate because they change their meanings in different contexts, are temporarily treated as untranslatable. In further linguistic analysis of this catagory, it is hoped that appropriate English syntactic structures that transfer the exact meaning of these words will be found.

Specific Word Indicators. Con-adjectival-adverbial, connominal, and special (H) classes of words (e.g., 地, 得, 关於, and 的) have specific tags, so that no specific Chicodes are needed to identify them.

Person. Nominals and verbals use person tags for English person verb inflections in reference to first, second, and third person, since Chinese has no verb person inflection.

Tags of specific interest will be referred to and explained in detail in the discussion of various word classes and of the utilization of information tags.

3.2 NOMINALS

The nominals have ten information tags, shown in Table 3-3, that include Chinese and English grammatic tags. In the study of the nominals, several problems were considered. The first problem involves the fact that differentiation must be made among pure nominals, verbal/nominals (V/N), and adjectival/nominals (A/N). Analyses of nominals in terms of semantic differences were undertaken, and nominals were grouped into types for proper translation in their various relationships with other words. Studies were made concerning: (1) nominals that, when preceded by 有 (9Q11)* and followed by 的 (VF11), can be changed into English adjectival forms, and (2) nominals that, when followed by 上, can be changed into English adjectival or adverbial forms.

The problem of utilizing one-character words that have functions in specific word classes, but may also serve as family names and therefore require romanization, was also taken into account. Words such as 總統 (president) and 主席 (chairman), which can also be used as titles for proper names, were analyzed for the purpose of assigning appropriate subtags.

There is also the problem of categorizing nominals that function as tense indicators or as collocative phrase endings. Person, number, capitalization, and translatability of the nominal class were analyzed in the study of nominals.

3.2.1 Nominal Tag 1—Kind

The most common kinds of nominal ambiguities are V/N and A/N, which are different from pure nominals; pure nominals have only the nominal grammatic function. Two examples of pure nominals are:

1. 人民 (people)
2. 战争 (war).

A V/N is a nominal that can also function as a verb. Three examples of V/N's are:

1. 工作 (work)
2. 代表 (represent, representative)
3. 組織 (organize, organization).

A V/N is different from a pure verbal with an English equivalent noun form. For example, in 他的来到 (his coming), 来到 (come) is not a V/N, but a pure V that has an English noun form. Usually, 的 (HM) is not needed to make a V/N function as a nominal, while a pure V requires that HM precede it, e.g., 共产組織 (Communist organization).

An A/N is a nominal that sometimes functions as an adjectival. The following are examples of A/N's:

1. 快樂 (happy, happiness)
2. 困难 (difficult, difficulty)

---

* Codes in parentheses refer to Chicodes for the Chinese character.

3. 健康 (health, healthy).

When functioning as adjectivals, these words must usually be followed by HM or preceded by ADRR, such as 很 (very), 十分 (very), and 相当 (quite). Two examples are:

1. 困难的问题 (difficult problem)
2. 健康的生活 (healthy life).

However, when these words function as nouns, they need not be preceded by HM. For example:

1. 经济困难 (economic difficulty)
2. 心理健康 (psychological health).

A pure adjectival, however, must have HM preceding it to cause it to take noun form.

Although words such as 历史 (history) and 农业 (agriculture) have equivalent English adjectival forms (historic, agricultural), they are entered as nominals with adjectical forms. A rule causes them to take adjectival form when they precede a nominal without HM in between:

- $ND_1 + N_2 \rightarrow AJXXXX_1 + N_2$

3.2.2 Nominal Tag 2 — Type

The nominals are divided into eleven subclasses, as follows.

Pronominal (M). This subclass includes all pronouns, e.g., 你 (you), 我 (I), and 他 (he).

Country and Continent (C). This subclass includes all nominals that are proper names of countries and continents. Although these are classified as pure nominals, the adjectival forms are included in the English translation. Some examples are: 美国 (U.S., American) and 非洲 (Africa, African).

Idea (D). This subclass includes the designations of theory, point of view, concept, life, composition, etc.—names of nonconcrete things that are abstractions or generalizations, e.g., 社論 (editorial), 会議 (conference), and 反面 (opposite). This category of nominals sometimes utilizes the YD&S tag (tag 7).

Organization (O). This subclass includes the designations of agency, group, company, etc.—nominals that have a collective, active authority to perform specific activities. Some examples are: 代表团 (delegation), 党 (party), and 国家 (country).

Human (H). This subclass includes terms for human individuals. Some examples are: 夫人 (madam), 主席 (chairman), and 大使 (ambassador).

Human Collective (E). This subclass includes generalized names for groups of people or for nonspecific persons. The English translation for such words is usually plural in number. Discretion should be used in differentiating these nominals from human nominals, which are mainly singular in number. Human collective nominals do not use the address tag (tag 8). Some examples are: 人民 (people), 工人 (workers), 学生 (students), and 資本家 (capitalists).

Inanimate (I). This subclass is defined to include all things that are concrete and neither human nor animal, such as stones, trees, places with common names, radio stations, parts of the body, and buildings. Some examples are: 房子 (house), 衣服 (clothes), and 汽車 (automobile).

Beasts (B). This subclass includes all live animals, singular or collective. Some examples are: 家禽 (domestic fowls), 狗 (dogs), 羊 (sheep), 鳥 (birds), and 魚, (fish).

Place Names (P). This subclass includes all proper place names, excluding countries and continents. Some examples are: 柏林 (Berlin), 上海 (Shanghai), and 波士頓 (Boston).

Nominals of Time (T). This subclass includes special nominals that involve the concept of time. These function as tense indicators and utilize the definity tag (tag 9). Some examples are: 八月 (August), 今天 (today), 年 (year), 現在 (present), and 时候 (time).*

Collocative Terminals (K). This subclass includes nominals that are basically nominals, but are often used as collocative phrase endings. Some examples are: 方面 (aspect) and 一方面 (aspect).

3.2.3 Nominal Tag 3—Person

All nominals are in the third person except 你 (you), which is in the second person, and 我 (I) and 本人 (I), which are in the first person.

3.2.4 Nominal Tag 4—Number

The number tag is classified according to the English translation of the Chinese word.

3.2.5 Nominal Tag 5—Capitalization

Proper nominals use P for the capitalization tag; common nominals use 0.

---

* See Section 3.2.9 for specific discussion of the nominal of time.

3.2.6 Nominal Tag 6—Gender

Nominals of the subclasses denoting idea, organization, inanimate, continent, place name, beast, nominal of time, and collocative terminal usually use N (neuter) for their gender tag. Country names usually use F (feminine). Nominals that are concerned with human beings and are not specifically known as either feminine or masculine (M) are all considered to be masculine. For example, 学生 (students) and 人民 (people) use M in the gender tag, but 女孩子 (girl) and 妇女 (women) use F in the gender tag.

3.2.7 Nominal Tag 7—YD&S Tag

Some nominals in the subclass "idea" go inside the phrase 有 ... 的 and become adjectival. Some examples are:

1. 有勇气的 (brave)
2. 有意义的 (significant).

These idea nominals (勇气 and 意义) use Y in the YD&S tag, and the English adjectival forms (brave, significant) are included in the English translation. However, single character idea nominals such as 力 and 利, which also go into the YD phrase and become adjectival, need not have the YD tag, and the English adjectival forms need not be included in the English translation.

Some nominals followed by 上 become adjectival or adverbial, e.g., 政治上 (political, politically), and 詞彙学上 (lexicographical, lexicographically). These nominals (政治 and 詞彙学) use S in the YD&S tag, and the A forms are included in the English translation.

3.2.8 Nominal Tag 8—Address

Some human nominals can sometimes be used as the title of a proper name. For example:

1. 王先生 (Mr. Wang)
2. 毛主席 (Chairman Mao).

The human nominals (先生 and 主席) use A in tag 8.

3.2.9 Nominal Tag 9—Definity

This tag is used primarily for the nominals of time, which are words concerning the conception of time. A T is used as indicator in the nominal tag 2, and tag 9 is used to indicate the subclasses into which the nominals of time are divided. At present, there are six major subclasses, as follows.

Nominal of Time—Definite. The names of the twelve months of the year, the seven days of the week, the different seasons, the different periods of time of day and night, etc., are included in this subclass. A D is put in tag 9 as indicator. When these words are not preceded by other nouns of time or followed by either VS1 (是) or HM (的), they are treated as collocative phrases at a certain phase of the translation scheme. Some examples of such words are: 星期五 (Friday), 中午 (noon), 春天 (spring), and 十二月 (December).

Nominal of Time—Indefinite. Another group of nominals of time has the function of collocative termination. The nominals in this group are classified as indefinite nominals of time rather than as collocative terminals. If one of these words is preceded by a collocative initial, it functions as a terminal. If not, it is a regular nominal of time. An I in tag 9 is used as indicator. Some examples of such words are: 時候 (time), 之日 (day), and 的同時 (at the same time).

Nominal of Time—Numeral. A few words are used to measure time, and are classified as nominal of time numerals. Like other measurement words in the synonome class, when these words are preceded by a numeral the resulting phrase can be either a nominal or an adverbial phrase. Some examples are: 年 (year), 天 (day), and 星期 (week).

Nominal of Time—Relative. There is a group of words that indicate the relativity of time. In Chinese, words of this group are used to indicate the time of the sentence relative to the present. Such words are classified as relative nominals of time. An R is put in tag 9 as indicator. Since these words usually overlap with the adverbial class, they are classified as $\mathrm{AD/NT\overset{9}{R}}$. The adverbial tag 5 indicates tense (see Section 3.3). When these words precede any other nominal of time, a nominal of time phrase is made and the tense tag is kept. Later the tense tag is reduplicated in the verbal tag 15 to indicate the correct tense inflection of the English verb form. When these words precede HM, they take noun form and a regular noun phrase is made. If they precede any other word, they take the adverbial form. The rules generated are:

1. $\mathrm{AD_1/NT\overset{9}{R}_2 + NT\overset{9}{D} \rightarrow \underline{NT\overset{9}{D}_2} + of + AD_1}$
2. $\mathrm{AD_1/NT\overset{9}{R} + HM + NX_2 \rightarrow NX_2 + of + NT\overset{9}{R}_1}$
3. $\mathrm{NX + AD/\overset{9}{NTR} \rightarrow NX + AD}$
4. $\mathrm{AD/\overset{9}{NTR} + NX \rightarrow AD + NX}$

Some examples of relative nominals of time are:

1. 去年 (last year)
2. 明天 (tomorrow)
3. 昨晚 (last night).

Nominal of Time Indicator. Certain words, such as 紀元(A.D.), 西元前(B.C.), and 公元 (A.D.), are used to indicate the time of a nominal of time phrase. The words are classified as nominal of time indicators. An N is put in tag 9 as indicator. The rule generated is:

- $\overset{9}{\text{NTI}} + NP_T \rightarrow NP_T + \overset{9}{\text{NTI}}$

Nominal of Time—O. The $\overset{9}{\text{NTO}}$ word (日) is different from other nominal of time numerals because of its irregular relationship to numerals. When this word occurs after a numeral and the numeral is preceded by a definite nominal of time, the numeral is to take the UQ form, and $\overset{9}{\text{NTO}}$ is to be deleted. An O in tag 9 is used to indicate this word.

3.2.10 Nominal Tag 10—Translatability

This tag refers to the translatability of the entry. A 0 indicates that it can be translated, and a U that it cannot be translated.

3.3 ADJECTIVAL ADVERBIALS

The adjectival-adverbial class, shown in Table 3-4, includes all Chinese morphemes that modify verbs or nouns. This group of modifiers is under the general heading of A, since some words can be used to modify both verbs and nouns, and word classes would otherwise have to be further subdivided into similar subtags.

3.3.1 Adjectival-Adverbial Tag 1

This tag is always A to distinguish it from other grammatical classes.

3.3.2 Adjectival-Adverbial Tag 2

This tag indicates the major divisions of the A class. Morphemes that are adjectivals, i.e., that can be used as attributes of nouns with or without the adjectival indicator 的(HM), or that can be attributes of verbs or verb phrases only when followed by the adverbial indicator 地(OA), are classified as subclass J. Examples of such adjectivals are:

1. 崇高 (lofty)
2. 緊張 (tense)
3. 合法 (legal)
4. 卑鄙 (abject)
5. 不得人心 (unpopular).

Morphemes that are usually used as attributes to verbs without being followed by OA, but that must be followed by HM when used as attributes of nouns, are classified as D. Some examples of members of this class are:

1. 充分 (adequately)
2. 起碼 (minimum)
3. 任意 (willfully)
4. 进一步 (further).

Since the linguistic analysis has not yet gone into the level that distinguishes a subordinate clause from a main clause in a complex sentence, all subordinate clause indicators such as 不但 (not only), 雖然 (although), 如果 (if), and 而且 (and) are temporarily included in this subclass.

Morphemes that need not have HM or OA to be attributes to either noun or verbs are classified as A. Examples of members of this class are:

1. 非法 (illegal)
2. 积极 (positive)
3. 集体 (collective)
4. 連續 (continuous)
5. 全面 (overall).

We do admit that the above classification depends a great deal on the subjective use of the classifier. It is hoped that the criterion of classification is always the structure most commonly used in modern Chinese (白話). According to the above classification, the following linguistic rules are generated:

1. $AA_1 + OA_2 + V_3 \rightarrow AD_1 + V_3$
2. $AJ_1 + OA_2 + V_3 \rightarrow AD_1 + V_3$
3. $AA_1 + HM_2 + N_3 \rightarrow AJ_1 + N_3$
4. $AD_1 + HM_2 + N_3 \rightarrow AJ_1 + N_3$
5. $AA + V \rightarrow AD + V$
6. $AA + N \rightarrow AJ + N$

There are some Chinese adjectivals whose English equivalents must go after the nouns they modify. In such cases, the adjectivals are classified as AP (adjectival postal) or AB (adjectival postal or adverbial). Some examples are:

1. 海外同胞 (compatriots abroad)
2. 成套的机器 (machines in sets)

3. 超额的收成 (harvest above norm).

When such classes appear, what is best and simplest for the longest match method and for machine operations is usually taken into consideration.

There is a group of adjectivals whose members have a very close relationship with certain connominals. These adjectivals are classified as special adjectivals (AS) with a completely different set of tags. Some examples of such special adjectivals are:

1. 远 (far)
2. 近 (near)
3. 一样 (same)
4. 一伙 (together)
5. 一起 (together).

All adjectivals and adverbials take noun form when preceded by HM and followed by a phrase segmentation indicator (KXX):

- HM + A + KXX → HM + N + KXX

3.3.3 Adjectival-Adverbial Tag 3

This tag indicates the degree of the adjectival or adverbial. In Chinese the degree indicator is a group of independent words, while in English the degree indicator is either a separate word or a suffix. Most Chinese adjectivals and adverbials are therefore in regular form in the dictionary. In context, when they are preceded by a degree indicator, a rule will combine the two Chinese words into one, as follows:

- $\mathrm{ADC}_1 + \mathrm{AJR}_2 \rightarrow \mathrm{AJC}_2$

From the English tags, the machine will be able to find the correct form for the English equivalent of the two Chinese words.

3.3.4 Adjectival-Adverbial Tag 4

This tag is used to indicate the quality of the adverb. Most adverbials are attributes only to verbals. A V or a K is used in tag 4. In Chinese, most adverbials precede the verbs they modify, and in such cases they can be used as verb indicators. In English, however, some adverbs usually go after or before the verbs they modify. The classifier is able to indicate this by choosing a V or a K. The following rules are generated for this purpose:

1. $ADXV_1 + V_2 \rightarrow V_2 + ADXV_1$
2. $ADXK_1 + V_2 \rightarrow ADXK_1 + V_2$

It is discovered that degree indicators (ADCR, ADSR) are the only adverbs that can modify verbs, adjectivals, and adverbs. Some examples are: 最喜欢 (like most) 更好的做法 (better way) and 做得更好 (do better).

There are certain adverbs that can only modify adjectives (ADRJ), e.g., 高度机密文件 (highly secretive document).

There are certain adverbials that do not influence the grammatical structure of the sentence. Such adverbials are called " independent adverbials." Because of the lack of time, the clause conjunctions are also temporarily put into this class. Some examples of such adverbials are:

1. 多年以来(for many years)
2. 但是 (but)
3. 在实际上 (actual)
4. 全体(all).

3.3.5 Adjectival-Adverbial Tag 5

There is a group of modifiers that is used in Chinese to indicate time. Tag 5 is used for such purpose. In machine operation, the adverbial tag 5 is moved into the verb tag 15, causing the English verb to have the correct English tense inflection. Some examples are:

1. 現在 (present)
2. 以前 (in the past)
3. 目前 (at present)
4. 当年 (in those days).

3.3.6 Adjectival-Adverbial Tag 6

There are certain important Chinese verbal indicators for which English does not have any equivalent single words. For the present, they are not translated, but as the linguistic system is refined, it is hoped that the correct translation may be decided on. Examples of such words are: 就, 都, and 才.

The major linguistic classes that overlap class A are A/N, AD/V, and AJ/V. The A/N's are introduced by the nominals. The AD/V's are Chinese words that function as verbs when they are not followed by a verb, but become attributes of verbs when they precede verbs in context.

Instead of classifying such words as VTE or VTD, it is much more convenient to classify them as AD/V. Some examples of such words are:

1. 努力 (diligently, work hard)
2. 繼續 (continuously, continue)
3. 分开 (separately, separate).

The following linguistic rules apply:

1. $AD/V + V \rightarrow AD + V$
2. $AD/V + \$\overline{V}^{*} \rightarrow V + \$\overline{V}$

Class AJ/V includes all words ending with 化 , such as 工业化 (industrialized) and 法西斯化(fascistize), and other words like 进步 (progressive) and 巩固 (solid).

3.4 VERBALS

In the course of this study, the verbal class has been found to be the most interesting and the most challenging, since it plays such an important role in sentence structure. Aside from its primary function as the main verb in a sentence, the Chinese verb alters its form depending on the phrase structure in which it is used. Several problems arise from this peculiarity of the Chinese verb:

1. There is no inflected form for conjugation with person, number, or tense. The tense and aspect can be specified by adverbial of time indicators and by other indicators, such as auxiliaries and converbals.

2. In general, there is no distinction in Chinese verbs between transitivity and intransitivity. With the exception of a limited number of verbs that do not take objects, most verbs can take objects or their equivalent. It is therefore necessary to distinguish the types of object and indirect object that the verb may take.

3. The verb plays an important part in influencing the translation of certain connominals (prepositions). There are several connominals whose translations change according to the verb used.

4. The bifunctional or even multifunctional nature of some verbs calls for different classifications for the same word.

---

[*] The symbol $\$\overline{}$ means "anything except...."

5. The verbal subtags concerning Chinese and English grammatical functions must both be retained in the linguistic processing for optimum phrase and sentence structure analysis. As the linguistic research progresses, more tags may have to be added for further refinement of the linguistic system.

3.4.1 Verbal Subtags

The verbals have 16 subtags, as shown in Table 3-5. Tags 1 through 11 are Chinese grammatic tags, and are included in the dictionary entry for each verb. Tags 12 through 16 are English grammatic tags, and are inserted through linguistic processing by machine operation.

Verbal Tag 1. This tag identifies the major grammatic verbal class (V).

Verbal Tag 2. This tag identifies the verbal types, of which there are five: transitive (T), intransitive (I), transitive and intransitive (B), special (S), and verbs of adverbial quality (W).

Verbal Tag 3. This tag identifies the object of the verb. The object includes the material that immediately follows the verb and/or that is grammatically important to sentence structure. The object might therefore include verbs and embedded sentences. When the object is in verb form, its English equivalent is usually in infinitive form.

Verbal Tag 4. This tag identifies the complement that follows the object. Because of its position, a word that is ordinarily called an indirect object is considered for machine purposes to be a complement.

Verbal 5. This tag is used to indicate a pretransitive, a type of connominal that functions as an indicator of the object before the verb, such as 把 . An example is: 請你把这本书給他 (please give him this book).

Tags 6 through 11 are called connominal tags. Several very common connominals change their translation according to individual verbs. For example, 对 can be translated seven ways and 向 can be translated six ways, depending on the verb. Seven kinds of connominals are therefore included so that the proper translation of the connominals can be indicated by the verb tags.

Verbal Tag 6. This tag is called the connominal-C tag. It deals with the connominal IC(对). This connominal can be translated seven ways:

1. Translate as "to": 我对他說(I say to him)
2. Translate as "concerning": 他对这件事情表示同意(he expresses agreement concerning this matter)

3. Translate as "in": 我对机器翻譯很感兴趣(I am interested very much in machine translation)

4. Translate as "with": 印度拿这个條件对中共妥协(India takes this condition to compromise with Red China)

5. Translate as "against": 美国对日本宣戦 (the U.S. declares war against Japan)

6. Translate as "from": 政府对人民征收所得税(the Government collects income tax from the people)

7. Translate as "at": 我们对敵人开枪 (We shoot at the enemy).

Verbal Tag 7. This tag is called the connominal-D tag. It deals with the connominal ID (向). This connominal can be translated six ways, depending on the verb:

1. Translate as "to": 我向他解釋(I explain to him)
2. Translate as "from": 我向他借一本书 (I borrow a book from him)
3. Translate as "against": 向苏聯备戦 (prepare war against Russia)
4. Translate as "toward": 他们向学校走去 (they walk toward the school)
5. Translate as "at": 資本家向他们的下属咆哮(capitalists roar at their subordinates)
6. Translate as "with": 我向他斡旋这一件事情(I conciliate this matter with him).

Verbal Tag 8. This tag is called the connominal-K tag. It deals with the connominal IK (为) in the postverb position, and it can be translated in two ways:

1. Translate as "into": 把中国改造为共产主义国家(reconstruct China into Communistic country)

2. Translate as "as": 我们从他为領袖 (we consider him as leader).

Verbal Tag 9. This tag is not at present used for Chinese grammatic information, but it is retained for further linguistic development of verb and connominal relationships.

Verbal Tag 10. This tag deals with connominals that have irregular verbal forms. An N is for connominal IN (成), which is translated as "into." An O is for connominal IO (作, 作为, 做), which is translated as "as." These are illustrated as follows:

1. IN translation in relation to verb: 共产党把农民組織成軍隊 (Communist Party organizes farmers into troops)

2. IO translation in relation to verb: 美国帝国主义把越南看作为殖民地 (U.S. imperialists see Vietnam as colony).

Verbal Tag 11. This tag concerns the use of connominals IA (同,和,与,跟), which are ordinarily used as conjunctions but become connominals when used with certain verbs.

Tags 12 through 16 concern the English functions of the verb. The English forms of the verb are produced during the linguistic loop operations, i.e., during machine operations and table lookup, which add to or refine the English tags during the passes that generate linguistic rule lookup.

Verbal Tag 12. This tag gives the main forms of the verb—main verb (V) and negative verb (Y).

Verbal Tag 13. This tag indicates person—first, second, or third—infinitive form of the verb (I), present participle form (P), past participle form (D), and auxiliary form (G).

Verbal Tag 14. This tag indicates number—singular (S) or plural (P).

Verbal Tag 15. This tag indicates tense—present (P), present perfect (E), past (A), past perfect (S), future (F), future perfect (U), or present progressive (R).

Verbal Tag 16. This tag indicates voice—active (A) or passive (P). As the linguistic analysis progresses, more form indicators may be added. A separate section on the English output is included in Appendix C.

Following is a discussion of the four types of verbs so far encountered.

3.4.2 Intransitives

There are four kinds of situations in which we consider the verb to be intransitive:

1. When a Chinese verb is used that truly takes no object, such as 来临 (arrive) and 爆发 (explode).

2. When a Chinese verb is used that is intransitive in the active voice, and whose English equivalent must be in the passive voice. For example, the verb 誕生 would be "be born" in English.

3. In Chinese the object is included in the verb, while the English equivalent is one word. For example, the verb 下雨 is actually composed of verb and object in Chinese, while the English equivalent is one word, "rain." To save computer operation and lookup, 下大雨 (rain heavily), 下小雨 (rain lightly), and 下毛毛雨 (drizzle) are also considered as single unit words by the longest match. As computer operation becomes more sophisticated, the division of these words for linguistic refinement will be considered. Other examples include such words as 耧地 (sow), 吃饭 (eat), 看书 (read), 走路 (walk), and 做事 (work).

4. In Chinese the object is included in the verb, and in English the translation includes the object. The reason for this classification is computer efficiency, as mentioned above. Examples of such words are:开会(hold meeting),領路(lead the way), and备戦(prepare for war).

On most occasions, the intransitive verb does not take verb subtags 3 through 11. However, sometimes the intransitive verb may take verb subtags 6 through 11, which concern the usage of connominals.

3.4.3 Transitives

We have the following types of transitive verbs in Chinese:

1. Verb VTA—verbs that take only a pure nominal or a noun phrase as object. If one of this class of verbs is not followed by a pure nominal or noun phrase, an operation to scan for noun possibilities is initiated. Some examples of these verbs are: 重建 (reconstruct) and 改变 (change):

- 我们要重建国家 (we shall reconstruct the country)
- 人民应該改变他们的观念(people should change their ideas).

2. Verb VTBW—verbs that must take pure nominals as their object, which may be followed by other objects. For machine purposes, the second object is considered to be complement, because of its position. Tag 4, the complement tag, therefore shows a W. Some examples of these verbs are: 給 (give), 送 (send), 寄 (mail), and 判处 (sentence):

- 郵差給了他一封信 (postman has given him a letter)
- 我寄他一本书 (I mail him a book)
- 法官判处他死刑 (judge sentences him death penalty).

3. Verb VTBI—verbs that must take pure nominals as their object, which may be followed by a verb whose English equivalent is in infinitive form. Some examples of such verbs are: 勒令 (impel), 領导 (lead), and 代表 (represent):

- 共党勒令学生退学 (Communist Party impels students to leave school)
- 中国政府領导人民反抗共产党(Chinese Government leads people to resist Communist Party)
- 聯合国代表人民反对侵略政策(U.N. represents people to oppose aggressive policy).

4. Verb VTCI—verbs that must take pure nominals as their object, which must be followed by a verb whose English equivalent is in infinitive form. Some examples of these verbs are: 使 (cause), 讓 (let), and 用 (use):

- 只会使人民觉悟起来(only will cause people to awake).

5. Verb VTD—verbs that may take pure nominals as their object or may be followed by a verb whose English equivalent is in infinitive form. Some examples of such verbs are: 準备 (prepare), 开始 (begin), and 进行 (proceed):

- 越共準备参加大規模戦争(Viet Cong prepares to participate in large scale war)
- 农民开始种稻 (farmers begin to plant rice)
- 我们进行反共运动(we proceed anti-Communist movement).

6. Verb VTE—verbs that must be followed by a verb whose English equivalent is in infinitive form. Some examples of such verbs are: 企图 (attempt) and 立意、(be determined):

- 中共企图侵畧东南亚(Chinese Communists attempt to invade Southeast Asia)
- 我们立意完成这件工作(we are determined to complete this work).

7. Verb VTF—verbs that can take either pure nominals or embedded sentences as their object. Examples of these verbs are: 証明 (prove) and 声明 (declare):

- 律師証明了他的过失 (the lawyer has proved his error)
- 这件事情証明他的观念是錯的(This matter proves that his idea is wrong.)
- 共产党声明将於月終完成一九六六年生产計劃 (Communist Party declares that it will complete 1966 production plans in the end of the month).

8. Verb VTG—verbs that must take embedded sentences as their object. Examples of these verbs are: 宣稱 (declare), 說 (say), and 暗喜 (be secretly glad):

- 中共宣稱将要进攻越南(Chinese Communists declare that they will attack Vietnam)
- 周恩来同志說「我们将尽一切力量援助越南同胞」(Comrade Chou En-lai says, "we shall exert all strength to help Vietnamese compatriots")
- 我们暗喜这一件事情快要成功了(we are secretly glad that this matter will succeed).

3.4.4 Special Verbs

At present, we classify the two special verbs 是 (MV11) and 有 (9Q11) as VS1 and VS2 respectively. Aside from its regular function as copula (the verb "to be"), 是, which is used mainly for emphasis and is untranslatable, very often goes freely either before or after the main verb. 是 is frequently used to introduce a relative clause.

The verb 有 functions in different ways, depending on the position it takes in a given sentence.

1. Immediately preceded by N or NP, 有 is translated as part of the verb "to have":

- 他有一本书 (he has a book)
- 中国人民有一个伟大的领袖(Chinese people have a great leader).

2. Immediately preceded by IP (connominal phrase), by LP (collocative phrase), by a comma, or by a period, 有 is translated as "there is (are)," and sometimes a "that" is inserted before the next verb, if any:

- After LP: 在中国境内有許多种族 (there are many races in the territory of China)
- After IP: 关於这个问题有各种意見 (concerning this problem there are various kinds of opinions)
- After a comma or a period: 有許多人喜欢游泳 (there are many people [that] like swimming].

3. 有 functions as an an adjective maker when it forms with a "discontinuent" structure, i.e., 有 ... 的 with an idea noun inserted between:

- 有秩序的行动 (orderly movement)
- 有勇气的青年 (brave youth).

3.4.5 VW—Verbs of Adverbial Quality

There is a group of Chinese words often used to express the state of the verbs immediately preceding them. Words of this group also serve very conveniently as verbal indicators. Grammatically, their function resembles that of adverbials. However, there is a grammatic gap between these Chinese words and their English equivalents. In Chinese they are adverbials, but their English equivalents are verbs followed by a participle or by an infinitive. For example, 不完 in the sequence 做不完 is an adverbial modifying the verb 做 (do). The correct translation of the phrase is "be unable to finish doing." ("Be unable to finish" is chosen for 不完 instead of "cannot finish" because "can" cannot have the future tense form 将要做不完, "shall be unable to finish doing.") Other examples of such Chinese words and their English translation are:

1. 做成功 (succeed in doing)
2. 做下去 (continue doing)
3. 做得到 (be able to do).

Both classification and linguistic rules are used to bring about the shift of grammatic classes in the simplest way for the machine.

These Chinese words are classified as VW instead of as ADXXXX because sixteen tag spaces are needed for the shift to verbal. Tag 3 is used to indicate the correct English verb form of the English equivalent of the Chinese verb preceding VW in the Chinese word sequence. This is decided by looking at the English translation of the VW and the English usage. The other tags, 4 through 16, are left blank.

When a VW is preceded by a verbal in context, a linguistic rule will cause these words to reorder, causing the VW to precede the verbal. Thus, the two words are rearranged according to the English word order. Tag 3 of the VW is scanned to find the indicator of the correct form for the verbal, and a proper indicator is accordingly moved into the verbal tag 13. Grammatically, the verbal is still the main verb; the Chinese tags of the verbal tags 2 to 11 are therefore moved into the second to eleventh position of the VW, and the verbal is ignored. The English equivalent of the VW will have the proper conjugation, tense, and voice when it comes to the pass in which subject and verb are found and connected. An example of a linguistic rule is:

- VXXXXXXXXXXX-----$_1$ + VW3-------------$_2$ → VW3-------------$_2$ + VXXXXXXXXXXX-----$_1$
  → VW3-------------$_2$ + VXXXXXXXXXXX-P---$_1$
  → VXXXXXXXXXXX-----$_2$
  + VXXXXXXXXXXX-P---$_1$

Some of these words are regular verbals when they are preceded by nominals or verbal indicators such as E, G, or AD. For example:

1. 他成功了 (he has succeeded)
2. 我们得到了胜利 (we have obtained victory).

In such cases, another set of regular tags is made in the dictionary entry with its proper translation. When the above situations arise in context, a linguistic rule will cause the machine to choose the proper set of verbal tags and translate accordingly.

3.5 ARITHMATES

Arithmates (see Table 3-6) are words that are commonly known as demonstratives. This class has two tags: tag 1 indicates the major class (R), and tag 2 indicates the number—singular (S) or plural (P). Examples of arithmates include: 这 (this), 这些 (these), 那 (that), 有些 (some), 一些 (some), and 各 (various). In the relationship between arithmates and nouns (synonomes) the arithmate changes the number tag of those classes. For example:

1. 这些小事 (these small matters)
2. 那人 (that person)
3. 那种植物 (that type of plant).

3.6 SYNONOMES

Synonomes (see Table 3-6) are words that are commonly known as classifiers. This class has two tags: tag 1 indicates the major class (S), and tag 2 indicates the translatability of a word (T or U). Some examples of translatable synonomes are:

1. 班 (class)
2. 杯 (cup)
3. 撮 (batch)
4. 噸 (ton).

Some examples of untranslatable synonomes are:

1. 个
2. 隻
3. 件
4. 辆.

Synonomes include measurement words such as 尺 (foot), 畝 (mou), and 斤 (catty).

After the processing of rules, tag 2 will change to indicate the singular (S) or plural (P) form of the English.

3.7 NUMERALS

Numerals (see Table 3-6) are number words that have two tags: tag 1 indicates the major class (U), and tag 2 indicates the singular (S) or plural (P) form. Some examples of numerals are:

1. 一 (one)
2. 二 (two)

3. 三 (three)
4. 十 (ten)
5. 百 (hundred
6. 千 (thousand)
7. 亿 (hundred million)
8. 百万 (million)
9. 零 (zero).

After the processing of rules, the English tags can become one of the following five forms: Arabic (UA), irregular (UI), UT in relation with OE, UQ in relation with OG, and OH. Illustrations of how these are converted to English grammatic tags are as follows:

1. UP + UP + UP + UP → UA + UA + UA + UA (5,432)
2. OH_1 (第) + UP_2 + UP_3 → UI_2 + UQ_3 (thirty-sixth—第三十六)
3. UP + UP (万) → UT + UI (twenty thousand—二万).

These various English forms are necessary because the change from Chinese to English numbers is not uniform in each case. A further study of this will be made so that all possible number combinations in Chinese will be correctly translated.

3.8 CONJUNCTIONS

At present, conjunctions (see Table 3-6) have only two tags: tag 1 identifies the major class (J), and tag 2 indicates equal conjunction (C). Examples of conjunctions that have been analyzed are:

1. 和 (and)
2. 或者 (or)
3. 同 (and).

3.9 AUXILIARIES

Auxiliary verbs (see Table 3-6) include such words as 应該要千万 (must) and 必, 必須 (should). In the processing of rules during the verb phrase making phase, the auxiliaries will help determine the verb and its proper form. At present, tag 2 serves to identify the particular group of auxiliary verbs. It is expected that the auxiliaries will be conjugated in future work.

3.10 CON-ADJECTIVAL-ADVERBIALS

There are a few words whose functions are unique. It is not convenient to include them in or organize them into linguistic classes. Therefore, a class with an initial tag O, the con-adjectival-adverbial class, is created for these words (see Table 3-6). Each word with the initial O tag has an alphabet letter as the second tag, so that instead of being referred to by Chicode, which complicates the machine operation, these words are referred to by unique tag words. All these words have specific rules to go with them. As the linguistic analysis continues, this class is expected to increase.

3.10.1 OA (地)

Since 地 is used much more as an adverbial marker than as a noun in political texts, it is temporarily classified as OA. When this word is used as a noun in words such as 平地 (level land), 高原地 (plateau), and 耕作地 (farming land), the longest match method is used to enter the words as one dictionary entry. In the present linguistic analysis, OA is considered to be an adverbial marker. Some examples of rules concerning this word are:

1. $\text{AJXXXX}_1 + \text{OA}_2 \rightarrow \text{ADXVXX}_1$
2. $\text{AXXXXX}_1 + \text{OA}_2 \rightarrow \text{ADXXXX}_1$

For example:

1. 快樂地 (happily)
2. 坚貞不屈地 (indomitably).

When we progress to the study of scientific texts, this word will be considered as an N/O ambiguity.

3.10.2 OB (得)

Word OB overlaps with the GX class. When it is preceded by a pure noun and followed by a verb or verb ambiguity, it is considered to be a GX. If it is preceded by a verb or verb ambiguity and followed by an A and a KXX, it is considered to be an adverbial marker. Rules and examples concerning this word are:

1. $\text{N}_1 + \text{OB/GX}_2 + \text{V/N}_3 \rightarrow \text{N}_1 + \text{GX}_2 + \text{V}_3$: 我们得工作 (we must work)
2. $\text{V/N}_1 + \text{OB/GX}_2 + \text{AXXXXX}_3 + \text{KXX} \rightarrow \text{V}_1 + \text{ADXXXX}_3 + \text{KXX}$: 工作得很順利 (work very smoothly).

Words such as 得到 (obtain), 得失 (gain and loss), and 得罪 (offend) are considered as longest matches.

3.10.3 OC (反)

This word is considered to be an adjectival marker. When it precedes any noun, it makes the whole phrase adjectival. The whole phrase is masked and a regular adjectival tag word is generated for linguistic operations. A rule concerning this word is:

- $\mathrm{OC_1 + NXXXXXXXXX_2 \rightarrow \underline{OC_1 + NXXXXXXXXX_2} + AJX00X}$

For example:

1. 反人民 (anti-people)
2. 反共产党 (anti-Communist party).

3.10.4 OD (次,回,倍,趟)

There is a group of words whose members are all translated as "time" and function as constituents of adverbial phrases when preceded by numerals. These words are classified as OD. In one of the linguistic passes, when OD is preceded by a numeral, both the numeral and the OD are masked and a new adverbial tag word is generated for linguistic operations. An example of a rule generated is:

- $\mathrm{UP_1 + OD_2 \rightarrow \underline{UP_1 + OP_2} + ADRI0X}$

For example:

1. 三次 (three times)
2. 一回 (one time).

3.10.5 OE (成)

This word overlaps with the connominal class in linguistic classification. When it is preceded by anything other than a numeral, it is considered to be in the connominal class. An example of an ambiguity rule is:

- $\mathrm{\$\overline{UX} + OE/IXXXX \rightarrow \$\overline{UX} + IXXXX}$

When OE is preceded by a numeral, it causes the numeral to take the UT form; both words are masked and a regular adverbial tag word is generated. An example of a rule is:

- $\mathrm{UP_1 + OE/IXXXX_2 \rightarrow \underline{UT_1 + OE_2} + ADRI0X}$

For example:

- 八成 (eighty percent).

3.10.6 OF (百分之)

When this word occurs, the numeral following it will be moved into the front position. If the resulting phrase precedes an HM or a nominal, "of" is added and a regular adjectival tag word is generated for linguistic purposes. If it precedes a phrase indicator (K) or a verb or verb phrase, an adverbial tag word is generated for the purpose. Examples of rules are:

1. $\mathrm{OF}_1 + \mathrm{UP}_2 + \mathrm{HM}_3 \rightarrow \underline{\mathrm{UI}_2 + \mathrm{OF}_1 + \mathrm{of}} + \mathrm{AJRXXX}$
2. $\mathrm{OF}_1 + \mathrm{UP}_2 + \mathrm{KXX}_3 \rightarrow \underline{\mathrm{UI}_2 + \mathrm{OF}_1} + \mathrm{ADRI0X} + \mathrm{KXX}_3$

For example:

- 百分之卅 (thirty percent).

3.10.7 OG (分之)

When this word is preceded and followed by numerals, the first numeral takes the UG form and the second takes the UI form. The order is then changed. An example of a rule is:

- $\mathrm{UP}_1 + \mathrm{OG}_2 + \mathrm{UP}_3 \rightarrow \mathrm{UI}_3 + \mathrm{UQ}_1$

For example:

- 五分之二 (two fifth).

It is regretted that "fifth" (UQ) is not pluralized. It is hoped that this will be corrected after more operations and rules relevant to this situation are added. At present, there is no pluralized entry or table of UQ forms.

3.10.8 OH (第)

The numeral following this word takes UQ form. An adjectival tag word is generated if it precedes an HM or a nominal. An adverbial tag word is generated if it precedes a phrase indicator (K) or verb or verb phrase. An example of a rule is:

- $\mathrm{OH}_1 + \mathrm{UX}_2 + \mathrm{HM}_3 \rightarrow \underline{\mathrm{UQ}_2} + \mathrm{AJRXXX} + \mathrm{HM}_3$

For example:

- 第二 (second).

3.11 PUNCTUATION

Punctuation marks (see Table 3-6) play an important part in determining the end of a structure, whether it be a phrase, a sentence, a series of words, etc. Some punctuation marks generate rules to indicate phrase segmentation. Tag 1 indicates the major class (P) and tag 2 indicates the particular punctuation mark. For example:

1. PC (,) → KCI + PC + KCT
2. PP (.) → KPT + PP + KPI
3. PE (!) → KPT + PP + KPI

The following gives the general investigation of punctuation marks:

1. ! . ? — terminal markers indicating end of sentence
2. " " — for titles, remarks, idioms, quotations
3. « » — for book titles
4. ( ) — for additional clarifications of a previous phrase or clause; additions
5. ...— for passages omitted
6. : — for beginning of remark or statement
7. , — for five functions:

- Separating adverb independent or adverbial phrases: 其中,我们发現了很多錯誤 (among these, we have discovered many errors)
- Separating clauses: 有的条件比較好,有的条件很差 (some conditions are comparatively better, some conditions are very poor)
- Separating noun phrase or clause from main verb: 我们从农业技术改革工作的实践中,总結了这样基本經驗 (we from realization of work of agricultural technique reform, have resulted in this kind of basic experience)
- Separating main clause from phrases: 农业技术改革必須有明确的目的,讲求经济效果 (reform of agricultural techniques must have clear aim, desiring economic effect)
- Separating a series of verb-object phrases: 我国人民举起了... 旗帜,发揚了...精神已经戰胜了... 灾害以及...经济困难(people of our country have raised ... flag, have developed ... spirit, have won over ... disaster ... and economic difficulty)

8. 、— for indicating series:

- Can take the place of "and" between two series: 重要原料、材料 (important raw materials, materials)
- Among series of three or more: 小麦、麥、棉花等等重要产品的生产 (wheat, oat, cotton etc. important products (HM) production → production of important products of wheat, oat, cotton, etc.)

9. ; — Sometimes used in place of comma.

3.12 SPECIAL WORDS—H CLASS

A special word class (see Table 3-6) was created for those words that do not belong to any specific functional word class but are nevertheless grammatically important. Tag 1 identifies the major class (H) and tag 2 identifies the specific word.

的 is given the tags HM. This word has been discussed in many papers, notably in great detail by the University of California.[7] HM functions as an important syntactic marker for the various forms of noun phrases we have analyzed.

们 is given the tags HN. This word serves to pluralize NH and NM nominals. A linguistic rule is generated in the linguistic processing operation:

- $NMXXXXXXXX_1 + HN_2 \rightarrow NMXPXXXXXX_1$

之 is given the tags HZ. This word has functions similar to those of 的.

3.13 CONNOMINALS

The words that introduce adverbial phrases but seldom need adverbial phrase ending words are classified as connominals (see Table 3-7).

3.13.1 Connominal Position Tags

The connominal class has five position tags, as follows.

Connominal Tag 1. An I in the tag 1 position is used to distinguish this class from other classes.

Connominal Tag 2. The tag 2 position is used for alphabets so that each group of similar connominals can be easily referred to without looking up Chicodes, and thus without extra machine operation.

Connominal Tag 3. The tag 3 position is used to indicate whether special lockup for certain words is to be made to decide the correct translation for a connominal that has more than one English equivalent. The letter R is used for regular connominals, which have only one translation or whose translations can be decided by linguistic rules. The letter I is an indicator for those connominals whose correct translation requires looking into the verb tags. An A is used for connominals whose translation is found after a lookup for special adjectives. The letter C, indicating that a comparative structure is involved, is for connominals.

Connominal Tag 4. This is a position tag that indicates the relative position of the connominal to the word concerned, so that the correct set of tags of another class can be consulted to find the correct equivalent.

Connominal Tag 5. This tag is used to indicate the major structure of the phrase following the connominal involved. The letter N stands for the noun phrase; the phrase following the connominal is usually a noun phrase. The letter V indicates that the phrase following can be either a noun phrase or a verb with its object. This tag is useful in locating relative clauses.

3.13.2 Connominal Tag Words

Since there is great difficulty in deciding the correct translation and the correct function of the group of words we call connominals, the following is only part of this group, concerning which some analysis has been made. The analysis is by no means complete. It is hoped that the list and the information tags will be increased in the near future.

IAAAN (同,跟,和,与). This set of connominal tag words is used for the irregular tags of the four Chinese words indicated. The regular function is conjunctional. Three irregular translations have been found:

1. Translated as "with": 我同他在一起 (I am together with him)
2. Translated as "as": 这个与那个一样 (this is same as that)
3. Translated as "from": 这个与那个不同 (this is different from that).

IBRRN (替,代). Since these two words are usually followed by another verb that seems to carry more grammatical importance and more meaning, they are classified as connominals introducing adverbial phrases instead of as VTBI verbals, which are followed by an object and an infinitive complement. The approximate English translation found is "for":

1. 我替他担心 (I am concerned for him)
2. 我会代你拿来 (he will bring for you).

ICIRN (对). Since this word is used in many different ways, much study of it has been done. It has been decided that the best way to handle its many translations is to include the information in the verb tags, since it changes its translation according to different verbals (see Section 3.4). It is one of the connominals that is included in the following structure causing reordering and insertion of words:

- $N_1 + IC_2 + N_3 + HM_4 + N_5 \rightarrow N_1$ (possessive) $+ N_5 +$ of $+ N_3$

Some examples are:

1. 他对历史的研究 (his study of history)
2. 我对他的意見 (my opinion of him).

IDIRN (向). We have also found that the translation of this word varies with the verbal. This information is also included in the verbal tags (see Section 3.4). In the conditions when it is followed by a collocative terminal or when the seventh tag of the following verbal tag word contains a zero, it is considered to be a collocative. For example:

- 他们向南方遷移 (they move toward the south).

IERBN (给). Besides being a VTBW verbal, this word is also considered to be a connominal. When it is preceded by a VTBW verb, it is a connominal introducing the indirect object and translated as "to." When it precedes a transitive verb, it is translated in two ways: (1) as "for" if the transitive verb is followed by an object, and (2) as "by" if the transitive is not followed by an object. In the latter case, the transitive verb will be changed into the passive voice form. For example:

1. Translated as "to": 我借给他一本书 (I lend to him a book)
2. Translated as "for": 这件事给我们带来了一些困难 (this matter has brought some difficulties for us)
3. Translated as "by": 这个東西给他找到了 (this thing has been found by him).

IFRRN (把, 将). When 将 (TN11) is followed immediately by a nominal, its function is the same as that of 把 (HJ11). The sole function of this class of connominals is the move of the direct object before the transitive verb. They are also called pretransitives. Sometimes 对 (YN11) and 向 (VA11) also have the same function. In such cases, the function is indicated in the verbal tag word. Some examples of members of the above class are:

1. 他把这件事忘了 (he has forgotten this matter)
2. 他将这件東西丢了 (he has lost this thing).

IGRIN (对于). At present, this word is translated as "concerning," but more study is expected to be done on it. An example is:

- 对于这件事，他没有发表意見 (concerning this matter, he has not declared opinion).

IHRIN (根据, 按, 按照, 就, 照,). This group is translated as "according to." Both IG and IH can introduce phrases that end with a member of the collocative terminal F class. In such cases, the collocative terminal overrides the connominal.

IIRRN (由). This word also overlaps with the collocative class. When it is followed by animate nominals (human collective, pronominal, etc.) it functions as connominal and is translated as "by"; if the following transitive verb does not have an object after it, the transitive is changed into passive voice. If the transitive verb has an object following, then it is translated as "let" and a set of its irregular tags as VTBU is taken. Otherwise, it is considered to be a collocative and translated as "from." For example:

1. Translated as "by": 这件事由他们处理 (this matter is managed by them)
2. Translated as "let": 我们决定由他们处理这件事 (we decide that we let them manage this matter).

IJRIV (由于). This word can be followed either by a clause or by a nominal phrase. When it is followed by a clause it is considered to function as a conjunction introducing an adverbial clause. Because of lack of time, it is classified as ADRI0T when it functions as an adverbial clause introducer. When it is followed by a nominal phrase, it is a connominal and is translated as "because of."

IKIBV (为). The translation of this connominal depends on its relative position and on the information carried in the verbal tags. When it precedes the verb, it is translated as "for." In such a position it can introduce a nominal phrase or a phrase made of a verbal and its object. In the latter case, the verbal is changed into the present participle form. When it is preceded by a verbal, it is translated in two ways. This is indicated in the eighth tag of the verbal tag word. Some examples are:

1. Translated as "for": 我们为你高興 (we are happy for you); 我们为反抗帝国主义而支持越南人民(we support Vietnamese people for resisting imperialism)
2. Translated as "into": 他们企图把中国变为殖民地 (they intend to change China into colony)
3. Translated as "as": 我们以此为我们的责任 (we consider this as our duty).

ILRRV (为了). This word always precedes the verbal to which it is attributive. It may introduce a nominal phrase or a phrase made of a verbal and its object. The verbal is changed into the present participle form. The connominal is also translated as "for."

INRPN (成,成为) and IORPN (做,作,作为). Tag 10 of the preceding verbal shows whether these two groups of words are connominals or verbs. If the tenth tag contains an N, IN is considered to be a connominal and is translated as "into." If the tenth tag contains an O, IO is considered to be a connominal and translated as "as." For example:

1. 他们把荒野变成稻田 (they change wilderness into rice fields)
2. 我们選他做主席 (we elect him as chairman).

IQRIN (特别是，尤其是). Any phrase that follows this group of connominals is considered to be an adverbial phrase and is masked in the linguistic process. Both words are translated as "especially":

1. 特别是这种产品 (especially this kind of product)
2. 尤其是在中国 (especially in China).

This is the only group of connominals that can be followed by any phrase other than a nominal.

IRCAN (比，比較). When 比較 (IJ11 DK15) precedes a nominal phrase, it is considered to be a connominal instead of ADCR0U. This group of connominals is a comparative structure indicator. Due to the lack of time, comparative sentence structure has not yet been analyzed. The future uses of this class have, however, been established. Aside from introducing an adverbial phrase, this class of connominals also causes the attributive, if there is one, to take the comparative form of the English equivalent.

ISAIN (像，如，如同，正如). The first three connominals of this class are sometimes related to special adjectives such as 一样 and 同样. The fifth tag of the special adjectives indicates the correct translation. When these connominals are not followed by a special adjective, they function in the same way as the remainder of this group and are translated as "such as," introducing a grammatically independent adverbial phrase. For example:

1. 我如他一样喜欢蘋果 (I like apples same as him)
2. 有些农业产品，如猪，羊等，(some agricultural products, such as pigs, sheep, etc.).

ITRRN (以). When this word appears in context, a processing tag will immediately search for IK (为). If IK is found, this word is considered to be a verb. If IK is not found, it is considered to a connominal and translated as "with." For example:

- 我应該以勇气对付这局面 (we should tackle this situation with courage).

IUCAN (没有，不及). This group is the negative of the IR group. It also indicates the comparative sentence structure pattern. For example:

1. 我跑得没有他快 (I do not run as fast as he)

2. 我不及他快 (I am not as fast as he).

IVAAN (离). This word becomes connominal if it precedes either of the two special adjectives 近 and 远 or a synonome phrase. It has two translations, and the fifth tags of these two adjectives indicate which of the forms to take. For example:

1. Translated as "to": 我家离基地很近 (my home is very near to the base)
2. Translated as "from": 这地方离文明世界很远 (this place is very far from civilized world); 此基地离台北有五里 (this base is five miles from Taipei).

3.14 CONVERBALS

Converbals play an important part in the analysis because, aside from solving verb ambiguities, they help to indicate the time and aspect of Chinese verbs, which have no inflection. Converbals also include indicators of negatives, present participles, indicators of passive voice, etc.

The converbal class takes four tags, as shown in Table 3-8. Tag 1 indicates the major class (E). Tag 2 indicates the types of converbals, such as tense indicators (T), negatives (N), present participle (P), passive voice indicators (A), and relative clause indicators (R). Tag 3 refers to tense indicators—initial (I) and terminal (T)—which designate whether the tense indicators go before or after the verb. Tag 3 indicates the tense for the negatives—present (P) and perfect (E). Tag 4 at this time refers only to tense indicators, and gives the tense—future (F), present progressive (R), and perfect (E).

Some examples of converbals are:

1. Tense and aspect indicators— 将要, 已经 (initial), 了, 过了 (terminal)
2. Negatives— 不, 没有
3. Present participle— 着
4. Passive voice indicator— 被
5. Relative clause indicator— 所.

3.15 COLLOCATIVES

The words concerning positions, time, and direction, which involve a discontinuous structure or introduce an adverbial phrase or clause, are classified as collocatives. As shown in Table 3-9, the collocatives are divided into two catagories according to their position relative to the phrase they enclose.

3.15.1 Collocative Initial (LI)

Words that indicate the beginning of a collocative structure are classified as collocative initial (LI). Some examples of such words are: 在 , 从 , 向 , and 沿. Words such as 除了, 自从, 当 , and 朝着, which introduce adverbial phrases, sometimes ending with a collocative word, are also included in this category.

3.15.2 Collocative Terminal (LT)

Collocatives that are at the end of a collocative structure are classified as collocative terminals (LT). Examples of such words are: 上 , 下 , and 之间 . Words that shown direction and position, such as 南方 and 外面, and idiomatic expressions such as 而論, 看来, and 来說, which may be the end of a phrase beginning with a connominal such as 照, or 根据, or with a collocative, 从 or 在 , are also included in this class.

3.15.3 Collocative Rules

To find the English equivalent for a collocative discontinuous structure, the collocative initial and the collocative terminal must be found and compared. For example, 在 ... 上 is translated as "on ...," and 在 ...南方 is translated as "to the south of ...." The collocative terminals and collocative initials are therefore further divided into various subclasses. Rules are made so that a certain class of collocative initial preceding a certain type of collocative terminal takes one of its two English forms, and vice versa. Some examples of the collocative rules are:

1. $\mathrm{LIA}_1 + \ldots + \mathrm{LTA}_2 \rightarrow \mathrm{LI1}_1 + \mathrm{LT2}_2 + \ldots.$
2. $\mathrm{LID}_1 + \ldots + \mathrm{LTA}_2 \rightarrow \mathrm{L1}_2 + \ldots.$

Each collocative terminal has an adverbial form. When a collocative terminal is immediately preceded by 在 (LID), it takes the adverbial form and LID is deleted. The rule generated is:

- $\mathrm{LID}_1 + \mathrm{LTX}_2 \rightarrow \mathrm{ADRV0T}_2$

For example:

1. 在南方 (in the south)
2. 在上面 (above).

When a collocative terminal is immediately preceded by any LI except LID, it takes the nominal form and the LI takes the regular form.

A rule generated is:

$\mathrm{LIX}_1 + \mathrm{LTX}_2 \rightarrow \mathrm{LI1}_1 + \mathrm{ND3S0N0000}_2$

For example:

1. 从南方 (from the south)
2. 向上面 (toward above).

If a collocative terminal is not preceded by a-phrase or by a phrase and a collocative initial, it takes the adjectival form. Some examples of the rules generated are:

1. $KXX_1 + LTX_2 + HM_3 \rightarrow KXX_1 + AXXXXX_2 + HM_3$
2. $KXX_1 + LTX_2 + KXX_3 \rightarrow KXX_1 + ADXXXX_2 + KXX_3$

For example:

1. 南方的人 (people in the south)
2. 后面的軍隊 (troops behind).

The collocative initial therefore has two forms in the dictionary entry: LI1 (zero) form and LI2 form. The collocative terminal has four English forms: LT1 form, used when the preceding collocative initial takes LI1 form; LT2 form, used when the collocative initial takes LI2 form; the noun form; and the A form (adjectival-adverbial form).

4. SYNTACTICAL ANALYSIS

The relationship between morphology and syntax and its problems and resolutions was analyzed in this course of study. This section includes discussions of noun phrases, verb phrases, collocative structures, connominal phrases, and noun phrase-relative clauses. In each analysis, the types of structure and their definitions are given, and the problems embodied in the syntactical analysis of the grammatic function of these structures are defined. The feasibility of the linguistic rules for application in machine translation is defined and illustrated through examples. Resolutions of the various problems of syntactical linguistic analysis are also discussed in this section.

Linking the verb to its subject in sentences is also discussed, and illustrations are given for the processing of this type of operation. To further enhance the details of this report, two sentences (Section 4.7), illustrated in art form, show the step by step linguistic processing.

All examples and illustrations given in this section are taken from the text material of "Hong Qi" magazines.[14]

4.1 NOUN PHRASES

A careful study of noun phrase structures and of their relationship to other phrases and structures at the sentence level has been made. In this study, the elements of the noun phrases were considered so that the proper recognition process for such phrases might be determined. In the investigation of all the possible combinations of noun phrases, it has been discovered that certain rules must be parsed before the search of the general table is made for noun phrases. Noun phrases have therefore been grouped into five categories, looked up sequentially. Over 150 noun phrase rules, which with programming will cover most of the possible noun phrase situations, have been established.

Briefly, the common types of noun phrases are as follows:

1. Noun phrases that require no reordering and include two nouns in succession or a noun with preceding modifiers (numerals, arithmates, synonomes, adjectives, etc.) For example:

- N + N: 美国人民 (American people)
- AJ + HM + N: 美丽的女子 (beautiful girl)
- U + S + AJ + HM + N: 一个进步的国家 (a progressive country).

2. Noun phrases that require some reordering and addition of English words or involve more than one noun. For example:
   - AP + HM + N → N + AP: 成套的机器 (machine in sets)
   - N_1 + HM + N_2 → N_2 + of + N_1: 中国的繁荣 (prosperity of China)
   - N + JC + N → N + JC + N: 中国和美国 (China and U.S.)
   - R + U + S + A + HM + N → R + U + S + of + N: 这三种笨重的机器 (these three kinds of massive machinery).
3. Noun phrases that include in the structure a collocative phrase (LP), i.e., noun phrase collocatives. For example:
   - LP + HM + N → NP_L: 在中国的人民 (people in China).

   The noun phrase collocative structure will be discussed in Section 4.3.
4. Noun phrases that include in the structure a connominal phrase (IP), i.e., noun phrase connominals. For example:
   - N + IC + N + HM + N → NP_I: 我们对和平的爱好 (our love of peace).
5. Noun phrase-relative clause structures. These will be discussed in Section 4.5, although they are considered to be noun phrase structures.

The noun phrase lookup operations have been tentatively divided into five categories. Basic linguistic operations for the noun phrases consist of reordering, inserting, altering, deleting, and masking.

4.1.1 Category 1

The first category is concerned with basic noun phrases that must be looked up before subsequent noun phrases. This process includes four rules.

1. ND3S0N000V + NXXXXXXXXX → $\underline{\text{ND3S0N000V}}$ + NXXXXX1XXX

清算　運動
(liquidate, liquidation)　(exercise, movement)　(liquidation movement)

The V in tag 10 indicates that the first noun is a noun form of an original verb. The function masks the first noun, and a 1 is inserted in tag 7 of the second noun to indicate that it is a noun phrase on the first level.

2. $\$\overline{\text{VTBWXXXXXXXXXXXXX}}$ + NCXXXXXXXX + NXXXXXXXXX → $\$\overline{\text{VTBWXXXXXXXXXXXXX}}$ + $\underline{\text{AJR00T}}$ + NXXXXX1XXX

保衛　美国
(protect, protection)　(U.S., American)　(people)　(protect American people)

In this situation, the rule does not apply if the first tag word is VTBW. When "anything except $[\$\overline{\quad}]$ VTBW" appears in this situation, the word will apply. The first noun alters all its tags to adjective form and is masked. The seventh tag of the second noun is altered to 1 to indicate that it is a noun phrase on the first level.

3. NHXX0XXXXX + NOXXXXXXXX → $\underline{\text{NXXP0XXPXX}}$ + NXXXXX1XXX

記者　工会
(reporter)　(union)　(reporters' union)

This rule applies when the first noun is human type and not capitalized and the second noun is organizational type. The function alters tag 4 of the first noun to plural form, and tag 8 indicates possessive case. The first noun is masked, and tag 7 of the second noun is altered to 1 to indicate that it is a noun phrase on the first level.

4. NHXXPXXXXX + NOXXXXXXXX → $\underline{\text{NXXXPXXPXX}}$ + NXXXXX1XXX

肯尼迪　政府
(Kennedy)　(government)　(Kennedy's government)

This rule applies when the first noun is human type and capitalized and the second noun is organization type. The function differs from example 3 in that proper names are not normally pluralized for the first noun. The function does not alter tag 4 of the first noun, but it indicates possessive case in tag 8. The first noun is then masked, and tag 7 of the second noun is altered to 1 to indicate that it is a noun phrase on the first level.

The above examples are only a few of the rules included in category 1, but they are used to illustrate the types of rules and the operations used for this category.

4.1.2 Category 2

The second category covers approximately 120 rules of noun phrases on the first level, and includes operations such as altering, inserting, or deleting tags, adding English words, and reordering or masking tag words. This category deals with general noun phrases, since the exceptional noun phrase first level lookup that is syntactically required is performed in the first category. The following are some examples of the general noun phrases covered in category 2.

1. $RS_1 + UP_2 + SU_3 + ADRAXX_4 + AJXXXX_5 + HM_6 + NXXXXXXXXX_7$

这 三 个 非常 精緻 的 花瓶

(this, these) (3, three) (very) (fine) (vase)

$\rightarrow \underline{RP_1 + UI_2 + ADRAXX_4 + AJXXXX_5} + NXXPXX1XXX_6$

(these three very fine vases)

By the longest match in the argument, when this rule applies, the function alters tag 2 of the arithmate (RS) to plural (RP). Tag 2 of the numeral (U) is designated irregular (I) so that it is translated as "three." Synonome untranslatable (SU) is deleted. Tag 4 of the noun is designated plural, and a 1 is inserted in tag 7 to indicate that the noun phrase is on the first level. Words 1, 2, 4, and 5 are masked, since these words are modifiers of the noun and are not essential in the operation for linking the verb to the main noun.

2. $US_1 + SU_2 + APXXXX_3 + HM_4 + NXXXXXXXXX_5 \rightarrow \underline{UI_1} + NXXSXX1XXX_5 + \underline{AJXXXX_3}$

一 位 海外 的 华侨

(one, an) (abroad) (overseas Chinese) (an overseas Chinese abroad)

When this rule applies, tag 2 of the numeral (U) is changed to irregular (I) so that it is translated as "an." Synonome untranslatable (SU) is deleted. The third tag word, APXXXX, which refers to a type of adjectical-adverbial class that should be placed after the noun, is reordered to follow the noun, and tag 2 is changed to adjective form (J). HM is deleted. Tag 4 of the noun is singular (S), since the numeral (U) is singular, and a 1 is inserted in tag 7 to indicate a first level noun phrase. Both UI and AJXXXX are masked, since they are irrevelant to subsequent operations.

3. $\mathrm{RS_1 + ST_2 + AJXXXX_3 + HM_4 + NXXXXXXXXX_5}$

那 种 合法 的 行动

(that, those) (kind, kinds) (legal, legally) (action)

$\rightarrow \underline{\mathrm{RS_1 + SS_2 + of + AJXXXX_3}} + \mathrm{NXXSXX1XXX}$

(that kind of legal action)

In this example, the arithmate singular (RS) alters tag 2 of synonome translatable to S. The English word "of" is added, and HM is deleted. Tag 4 of the noun is altered to S to indicate singular, and the seventh tag is altered to 1 to indicate that the noun phrase is a first level phrase. The first three tag words and the English word "of" are masked, since they are irrelevant for subsequent operations.

4. $\mathrm{KXX + NMXXXXXXXX_1 + HM_2 + NXXXXXXXXX_3 + KXX} \rightarrow \mathrm{KXX} + \underline{\mathrm{NMXXXXXPXX_1}}$

我 的 家

(I, me, we, us, etc.) (home)

$+ \mathrm{NXXXXX1XXX_3 + KXX}$

(my home)

This rule may be applied last when phrase indicators (KXX) are found before and after the noun phrase. Tag 8 of the first noun indicates possessive case, and HM is deleted. A 1 is inserted in tag 7 of the second noun to indicate the first level noun phrase. The first noun is masked.

4.1.3 Category 3

The third category is concerned with the processing of noun phrases on a second level. In category 2, noun phrases on the first level are processed; therefore, in category 3, noun phrases that include in the structure a connominal phrase (IP), noun phrases that include one or more first level noun phrases, and noun phrases that include one or more first level noun phrases and are separated by HM are processed. Operations include reordering and masking tag words or first level noun phrases, inserting English words, and altering, adding, or deleting subtags.

The following examples demonstrate the operations of noun phrases that include a connominal phrase in the structure.

$$1.\ \begin{bmatrix}\mathrm{KVT}\\ \mathrm{KCT}\end{bmatrix}^{*} + \mathrm{KII} + \mathrm{ICXXX} + \mathrm{KMX} + \mathrm{NXXXXXXXXXX}_1 + \mathrm{HM} + \mathrm{NXXXXXXXXXV}_2 + \begin{bmatrix}\mathrm{KCI}\\ \mathrm{KPT}\end{bmatrix}$$

对 国际工人运动 的 进攻

$$\rightarrow \begin{bmatrix}\mathrm{KVT}\\ \mathrm{KCT}\end{bmatrix} + \mathrm{NXXXXX2XXV}_2 + \underline{\mathrm{of} + \mathrm{NXXXXXXXOXX}_1} + \begin{bmatrix}\mathrm{KCI}\\ \mathrm{KPT}\end{bmatrix}$$

(attack of International Workers' Movement)

This rule deals with the ICXXX (对) connominal phrase within a noun phrase. Phrase indicators at the beginning and end of this phrase are necessary before this rule can apply. In this particular translation ICXXX is not translated, but the nouns are reordered. In the function, the first noun retains all its subtags but a 2 is inserted in tag 7 to indicate that it is a second level noun phrase. The objective case for the second noun is indicated by an O in tag 8. The English word "of" is inserted between the two nouns, and "of" and the second noun are masked for subsequent passes.

$$2.\ \mathrm{NXXXXXXXXXX}_1 + \mathrm{KII} + \mathrm{ICXXX}_2 + \mathrm{KMX} + \begin{bmatrix}\mathrm{N}\\ \mathrm{N + PS + N + JC + N}\\ \mathrm{N + JC + N}\end{bmatrix}_3 + \mathrm{HM}_4 + \mathrm{NXXXXXXXXXX}_5$$

美国統治集团 对 美国共产党 的 进一步迫害

$$+ \begin{bmatrix}\mathrm{KCI}\\ \mathrm{KPT}\\ \mathrm{KMX + LTX}\\ \mathrm{\$ + UP\ TO\ NEXT\ VXXXX0XXXXXXXXXXX}\end{bmatrix}_6 \rightarrow \underline{\mathrm{NXXXXXXPXX}_1}$$

, (American dominating group's

$$+ \mathrm{NXXXXX2XXX}_5 + \underline{\mathrm{of}} + \underline{\begin{bmatrix}\mathrm{NXXXXXXXOXX}\\ \mathrm{NXXXXXXXOXX} + \mathrm{PS} + \mathrm{N} \sim \overset{8}{\mathrm{O}} \sim {}^{10} + \mathrm{JC} + \mathrm{N} \sim \overset{8}{\mathrm{O}} \sim {}^{10}\\ \mathrm{NXXXXXXXOXX + JC + NXXXXXXXOXX}\end{bmatrix}_3}$$

further persection of U.S. Communist Party

$$+ \begin{bmatrix}\mathrm{KCI}\\ \mathrm{KPT}\\ \mathrm{KMX + LT}\\ \mathrm{\$ + UP\ TO\ NEXT\ VXXXX0XXXXXXXXXXX}\end{bmatrix}_6$$

,)

In this example, the tag words are again reordered and subtags are altered according to function. The above example actually shows 12 rules, since the tag words within the brackets show different possibilities. The second noun in the function is not masked, since all its tags are retained for subsequent passes. Ambiguities in the translation of the nouns are resolved by indication in tag 8 of the tag word. Again the ICXXX is not translated because it functions as a reorder indicator and is therefore untranslatable.

---

* The brackets include the alternative tag words, one of which is required.

$$3.\quad (\text{N})_1^{*} + \text{KII} + \text{IAAAN}_2 + \text{KMX} + \begin{bmatrix} \text{N} \\ \text{N} + \text{PS} + \text{N} + \text{JC} + \text{N} \\ \text{N} + \text{JC} + \text{N} \end{bmatrix}_3 + \text{HM}_4 + \text{NXXXXXXXXXV}_5$$

我　　　和　　　　共产党　　　　　的　　斗争

$$+ \begin{bmatrix} \text{KCI} \\ \text{KPT} \\ \text{KMX} + \text{LTX} \\ \$ + \text{UP TO NEXT VXXXXXXXXXX0XXXXX} \end{bmatrix} \rightarrow \underline{(\text{NXXXXXXXPXX})_1} + \text{NXXXXXX2XXX}_5$$

,　　　　　　　　　　　　　　(my　　　　　　struggle

$$+ \underline{\text{IAIXX}_2} + \begin{bmatrix} \text{NXXXXXXX0XX} \\ \text{N} \sim \overset{8}{\text{O}} \sim {}^{10} + \text{PS} + \text{N} \sim \overset{8}{\text{O}} \sim {}^{10} + \text{JC} + \text{N} \sim \overset{8}{\text{O}} \sim {}^{10} \\ \text{N} \sim \overset{8}{\text{O}} \sim {}^{10} + \text{JC} + \text{N} \sim \overset{8}{\text{O}} \sim {}^{10} \end{bmatrix}_3$$

with　　　　　　　　　　　Communists

$$+ \begin{bmatrix} \text{KCI} \\ \text{KPI} \\ \text{KMX} + \text{LT} \\ \$ + \text{UP TO NEXT VXXXXXXXXXX0XXXXX} \end{bmatrix}$$

,)

This rule concerns the IAXXX (同,跟,和) connominal phrase within a noun phrase. The first noun enclosed within the parentheses indicates that it need not be present for the rule to apply. The above example actually indicates 24 rules, since tag words within the brackets and parentheses show different possibilities. The connominal IAXXX selects the translation "with" by altering its tags to IA1XX (English grammatic tags), since the other two translations deal with special adjectives, and in this rule the other translations do not apply. Tag 8 of the nominals is changed to select the proper translation for the case. All tag words are masked except the fifth tag word, which indicates that the noun phrase is on the second level.

---

* Term in parentheses indicates that the tag word may or may not be used.

4. $\$\overline{\text{VTBWXXXXXXXXXXXXX}} + \text{NXXXXX1XXX}_1 + \text{NXXXXXXXXX}_2 \rightarrow \$\overline{\text{VTBWXXXXXXXXXXXXX}}$

各国人民 进步事业

$+ \quad \text{NXXXXX2XXX}_2 \quad + \underline{\text{of}} + \quad \underline{\text{NXXXXXXXXX}_1}$

(progressive enterprise of each country's people)

The above rule shows the reordering of first level noun phrases in relation with other nouns or first level noun phrases. The English word "of" and the second noun in the function are masked, and only the first tag word is left for subsequent operations.

4.1.4 Category 4

The fourth category deals with nouns or noun phrases that are separated by conjunctions or by series indicators. For this type of noun phrase, the subtag must be altered to indicate that the verb to be linked with it is to be pluralized.

1. $\text{NXXXXXXXXX}_1 + \text{PS}_2 + \text{NXXXXXXXXX}_3 \rightarrow \underline{\text{NXXXXXXXXX}_1 + \text{PS}_2} + \text{NXXXXXJXXX}_3$

和平 、 民主 (peace , democracy)

2. $\text{NXXXXXXXXX}_1 + \text{JC}_2 + \text{NXXXXXXXXX}_3 \rightarrow \underline{\text{NXXXXXXXXX}_1 + \text{JC}_2} + \text{NXXXXXJXXX}_3$

美国 和 中国 (U.S. and China)

Tag 7 of the second noun is indicated by a J, so that in the verb and subject linking pass a rule will be generated to pluralize the verb. A J is put in tag 7 so that tag 4 (the number tag for the tag word) retains its original form.

4.1.5 Category 5

The fifth category deals with noun phrases that include HD (等, 等等). The following examples mask the HD but retain the number tag for the noun for subsequent passes.

1. $\text{NXXXXXXXXX} + \text{HD} \rightarrow \text{NXXXXXXXXX} + \underline{\text{HD}}$

罗斯福肯尼迪 等等 (Roosevelt, Kennedy, etc.)

2. $\text{NXXXXXXXXX}_1 + \text{HD}_2 + \text{NXXXXXXXXX}_3 \rightarrow \underline{\text{NXXXXXXXXX}_1 + \text{HD}_2} + \text{NXXXXXJXXX}$

罗斯福,肯尼迪 等等 人 (Roosevelt, Kennedy, etc. people)

In the second example, a J is put in tag 7 of the second noun as an indicator so that the verb to be linked to this noun phrase will be pluralized.

The above five categories deal with noun phrases on the first and second levels. Noun phrase collocatives are resolved in the collocative structure pass (see Section 4.3). The noun phrase-relative clause, which includes a noun modified by a subordinate clause, is discussed in detail in Section 4.5.

4.2 VERB PHRASES

A Chinese verb phrase is a series of words that includes the main verb and/or one or more of the following elements: converbals, VW's, auxiliary verbs, and adverbs. These elements may or may not be present in the verb phrase structure. They aid in the specification of the English form for the Chinese verb (such as tense and aspect), which by itself lacks inflected forms. Some of the problems that have been found are:

1. The fact that converbals, auxiliary verbs, and adverbs may not be directly next to the verb in the processing sentence
2. Determining the word classes of individual Chinese words that influence the verb phrase with consideration to uniformity of subtags, translation, and their position relative to the verb
3. Determining the extent to which the converbals, auxiliary verbs, VW's, and adverbs aid in resolving verb/noun ambiguities
4. Determining what operations must be initiated before inserting the verb phrase initial and terminal indicators.

The following examples are given to show some of the verb phrase patterns discovered during the course of the analysis.

1. ADCVXX_1 + $\mathrm{VXXXXXXXXXXXXXXX}_2$ + ETTE_3 – KVI* + $\mathrm{VXXXXXXXXXXXXXEX}_2$

更加清楚地 暴露 了 (has/have exposed

+ ADCVXX_1 + KVT*

more clearly)

---

* KVI and KVT are verb phrase initial and terminal indicators.

2. ADRK0U$_1$ + ETIF$_2$ + VXXXXXXXXXXXXXXXX$_3$ – KVI + ADRK0U$_1$ + VXXXXXXXXXXXXXXFX$_3$

所以 要 迫害 (therefore will persecute)

+ KVT

3. ADRVAT$_1$ + ENIE$_2$ + EAI0$_3$ + VXXXXXXXXXXXXXXXX$_4$ – KVI + VTAXXXXXXXXXYXXSP$_4$

过去 没有 被 消灭 (had not been destroyed

+ ADRVAT$_1$ + KVT

in the past)

4. ADRV0T$_1$ + ETIF$_2$ + VXXXXXXXXXXXXXXXX$_3$ + ETT0$_4$ – KVI + VXXXXXXXXXXXXXXFX$_3$

終将 燃燒 起 (will burn

+ ADRV0T$_1$ + KVT

eventually)

5. ENIF$_1$ + VXXXXXXXXXXXXXXXX$_2$ – KVI + VXXXXXXXXXXXYXXFX$_2$ + KVT

决不 属于 (will not belong)

6. ADRV0T$_1$ + GN$_2$ + VTAXXXXXXXXXXXXXX$_3$ – KVI + VXXXXXXXXXXXXGXXX$_3$

一定 能够 取得 (can obtain

+ ADRV0T$_1$ + KVT

certainly)

7. ADRK0U$_1$ + VXXXXXXXXXXXXXXXX$_2$ + EPT0$_3$ – KVI + ADRK0U$_1$ + VXXXXXXXXXXXXXXRX$_2$

只不过 意味 着 (merely is implying)

+ KVT

8. ETIF$_1$ + ADXVXX$_2$ + VXXXXXXXXXXXXXXXX$_3$ – KVI + VXXXXXXXXXXXXXXFX$_3$

要 加紧 推行 (will promote

+ ADXVXX$_2$ + KVT

intensely)

9. GB_1 + ADXVXX_2 + $\mathrm{EIB0}_3$ + $\mathrm{VXXXXXXXXXXXXXXX}_4$ – KVI + GB_1

必須 不择手段地来 镇压 (must

+ $\mathrm{VXXXXXXXXXXXXXXX}_4$ + ADXVXX_2 + KVT

suppress unscrupulously)

10. ETIR_1 + ADXVXX_2 + $\mathrm{VXXXXXXXXXXXXXXX}_3$ – KVI + $\mathrm{VXXXXXXXXXXXXXXRX}_3$

正在 蓬勃地 发展 (are being developed

+ ADXVXX_2 + KVT

prosperously)

11. $\mathrm{VTEXXXXXXXXXXXXX}_1 + \mathrm{V}\underline{\mathrm{TAXXXXXXXXXXXXX}_2} - \mathrm{KVI} + \mathrm{V}\underline{\mathrm{TAXXXXXXXXXXXXX}_1}$

妄图 消灭 (attempt futilely

+ $\mathrm{VXXXXXXXXXXXXIXXX}_2$ + KVT

to destroy).

The verb phrase patterns shown above are but a few of the hundreds of patterns possible when we take into account the quality and position of adverbs, auxiliary verbs, converbals, and verbs. To include all forms of the verb phrases for the machine oriented linguistic analysis, specific operations to be accomplished for verb phrase lookup have been spelled out. The series of operations for verb phrase making is as follows.

1. Determining Phrase Initial and Terminal Indicators.

   - In matching verb phrase list, do not consider ADXXXX, i.e., jump over (or temporarily mask) ADXXXX. (This is done because ADXXXX can occur in any position in a verb phrase). For example, EXXX + ADXXXX + VXXXXXXXXXXXXXXX matches EXXX + VXXXXXXXXXXXXXXX.
   - Jump over (or temporarily mask) $\mathrm{V}\begin{bmatrix}\mathrm{B}\\\mathrm{T}\end{bmatrix}\begin{bmatrix}\mathrm{D}\\\mathrm{E}\end{bmatrix}\mathrm{XXXXXXXXXXXXX}$ if $\mathrm{V}\begin{bmatrix}\mathrm{B}\\\mathrm{T}\end{bmatrix}\begin{bmatrix}\mathrm{D}\\\mathrm{E}\end{bmatrix}\mathrm{XXXXXXXXXXXXX} + \mathrm{V\overline{W}XXXXXXXXXXXXXX}$. (This is done because $\mathrm{V}\begin{bmatrix}\mathrm{B}\\\mathrm{T}\end{bmatrix}\begin{bmatrix}\mathrm{D}\\\mathrm{E}\end{bmatrix}\mathrm{XXXXXXXXXXXXX}$ can occur within a verb phrase pattern in a series of more than one.) For example, $\mathrm{EXXX}_1 + \mathrm{V}\begin{bmatrix}\mathrm{B}\\\mathrm{T}\end{bmatrix}\begin{bmatrix}\mathrm{D}\\\mathrm{E}\end{bmatrix}\mathrm{XXXXXXXXXXXXX}_2$

$+\ V\begin{bmatrix}B\\T\end{bmatrix}\begin{bmatrix}D\\E\end{bmatrix}XXXXXXXXXXXXX_3 + VTAXXXXXXXXXXXXXX_4$ matches $EXXX_1$

$+\ VXXXXXXXXXXXXXXX_4$.

- If the above does not apply, treat $V\begin{bmatrix}B\\T\end{bmatrix}\begin{bmatrix}D\\E\end{bmatrix}XXXXXXXXXXXXX$ as a regular verb and match. For example, $EXXX_1 + V\begin{bmatrix}B\\T\end{bmatrix}\begin{bmatrix}D\\E\end{bmatrix}XXXXXXXXXXXXX_2 + VWXXXXXXXXXXXXXX_3$ matches $EXXX_1 + VXXXXXXXXXXXXXXX_2 + VWXXXXXXXXXXXXXX_3$.
- If match is found, insert KVI (verb phrase initial indicator) before verb phrase and insert KVT (verb phrase terminal indicator) after verb phrase.

2. Operations Within Phrase Indicators. Examination is made for the presence of certain occurrences within the verb phrase initial and terminal indicators. If these following occurrences do appear within the verb phrase indicators, operations are specified as follows:

- If VWXXXXXXXXXXXXXX is present:
    - If VW is not preceded by V, VW takes the verb regular form, which is specified in the lexicographic entry. The rule is:

$$\$\overline{VXXXXXXXXXXXXXXXX_1} + VWXXXXXXXXXXXXXXX/VXXXXXXXXXXXXXXXX_2$$

我们　　　　下去

$$-\ \$\overline{VXXXXXXXXXXXXXXXX_1} + VXXXXXXXXXXXXXXXX_2$$

(we　　　　go down)

-
    - If VW is preceded by V, VW reverses position with V, and tags 2 to 11 of V are inserted in VW, with a W inserted in tag 9 to indicate that it was originally a VW. Tag 13 of V indicates infinitive (I) or participle (P) form depending on the VW. The V is then masked. The rule is:

$$V\underline{XXXXXXXXXX}XXXXXX_1 + VW3XXXXXXXXXXXXXX_2 - V\underline{XXXXXXXWXXXXXXX}X_2$$

做　　　　下去　　　　(continue

$$+\ \underline{VXXXXXXXXXXXXXXXP_1}$$

doing)

-
    - If ENXX (ADXXXX) is present immediately before the verb (or with an adverb between), indicate negative form in tag 12 of V, jumping over ADXXXX. The rule is:

$\mathrm{ENXX_1 + (ADXXXX) + VXXXXXXXXXXXXXXX_2 \rightarrow VXXXXXXXXXXYXXXX}$

不　　　　喜 欢　　　　(do not like)

- If $\mathrm{V\begin{bmatrix}B\\T\end{bmatrix}\begin{bmatrix}D\\E\end{bmatrix}XXXXXXXXXXXXX}$ is present and followed by V, the $\mathrm{V\begin{bmatrix}B\\T\end{bmatrix}\begin{bmatrix}D\\E\end{bmatrix}\sim^{16}}$ takes the tags 2 through 11 from the V, with an E inserted in tag 9 to indicate that it was originally a $\mathrm{V\begin{bmatrix}B\\T\end{bmatrix}\begin{bmatrix}D\\E\end{bmatrix}}$. Tag 13 of the V is changed to I (infinitive form), and the V is then masked:

$$\mathrm{V\begin{bmatrix}B\\T\end{bmatrix}\begin{bmatrix}D\\E\end{bmatrix}XXXXXXXXXXXXX_1 + V\underbrace{XXXXXXXXXX}XXXX_2}$$

企 图　　　　消 灭

$$\mathrm{\rightarrow V\overbrace{XXXXXXXEXX}XXXXX_1 + \underline{VXXXXXXXXXXXIXXX_2}}$$

(attempt　　　　to destroy)

- If EXXX is present, put the qualities of EXXX into the verb and delete EXXX:

$\mathrm{ENXX + VXXXXXXXXXXXXXXX \rightarrow VXXXXXXXXXXYXXXX}$

$\mathrm{ETIF + VXXXXXXXXXXXXXXX \rightarrow VXXXXXXXXXXXXXFX}$

- If ADXXXX is present, the following sequential operations are necessary:
  - Put the fifth tag of $\mathrm{ADXX\bar{0}X}$ (tense indicator) into tag 15 of the verb. For example:

$\mathrm{ADXX\bar{0}X + (GX) + VXXXXXXXXXXXXXXX \rightarrow ADXX\bar{0}X + (GX) + VXXXXXXXXXXXXX\bar{0}X}$

  - Reorder ADXVXX and mask ADXXXX:

$\mathrm{ADXVXX_1 + (GX) + VXXXXXXXXXXXXXXX_2 \rightarrow (GX) + VXXXXXXXXXXXXXXX_2}$
$\mathrm{+ \underline{ADXVXX_1}}$

$\mathrm{ADXXXX \rightarrow \underline{ADXXXX}}$

- If GX is present, put a G in tag 13 of the verb and mask:

$\mathrm{GX + VXXXXXXXXXXXXXXX \rightarrow \underline{GX} + VXXXXXXXXXXXGXXX}$

It is hoped that the series of operations indicated above will cover all possible occurrences of varied forms of verb phrases and can therefore keep the verb phrase lookup table down to about 100 entries to determine positioning of phrase indicators. The series of operations within the phrase indicators will hopefully cover most operations to indicate both English grammatic tags for the verb and masking and reordering operations for subsequent passes to link the verb with the subject.

4.3 COLLOCATIVE WORDS AND STRUCTURES

Collocative structures include prepositional structures that are indicated by collocative initials or terminals, i.e., those words concerning positions, time, or direction (see Section 3.15). The collocative structure may involve a discontinuous structure such as 在 ...下 (under) or it may be an adverbial phrase or clause with a single collocative word indicator, such as 由 (from) ..., 除了 (besides).... The collocative structure may be isolated from the sentence by the recognition of the presence of a collocative initial and a collocative terminal in a discontinuous structure such as in 从这个时候起 (beginning from this time) or the collocative structure may be indicated by only a collocative initial, such as in 从台湾和台湾海峡 (from Taiwan and Taiwan strait), or by only a collocative terminal such as in 資本主义国家内 (in capitalist countries).

The problems concerning the analysis of the collocative structures can be divided into two levels: word and structure.

4.3.1 Collocative Words

First of all, collocative initials must be differentiated from collocative terminals. To determine how the translation is affected, a comparison has to be made between collocative initials and terminals when they occur in the collocative structure as initial and terminal words. When this comparison is fairly well analyzed, the collocatives are then divided into categories according to their function in the translation scheme. Thus, for example, 从 (from), which is classified as LIC, in the discontinuous structure of 从 (LIC)...看来 (LTD) has the translation: (blank)(LI1) "according to (LT1)..."; while in the discontinuous structure of 从 (LIC)... 来 (LTE), the translation is "from (LI2)...." The ambiguities of words such as 向 (L/I), 来 (L/V/E), and 里 (L/S), which have other functions than that of collocative, must be resolved through linguistic rules.

4.3.2 Collocative Structures

There is little difficulty concerning the problem of segmenting collocative structure(s) from the rest of the sentence if the structure contains both collocative initial and terminal words. If one collocative word is present without the other, the terminal or initial point of the structure must be found and indicated. When the initial and terminal points of the collocative structure are found, structure initial and terminal indicators are inserted. The content of the structure must be reordered, and significant tags are then indicated for proper translation. At the same time, the correct translation form for the collocative word(s) is determined. The collocative structure should be masked, since it is a modifying phrase or clause and is not needed for the verb-subject linking pass. When the structure is not at the beginning of a sentence and followed by a punctuation mark, the collocative structure should follow the verb and its object.

Collocative words have been discussed in Section 3.15. This section is therefore concerned with collocative structures, and collocative initial and terminal words are mentioned only if they are relevant to the clarification of collocative structure processing.

The collocative structure processing has been tentatively divided into the following steps:

1. Ambiguity resolving of collocatives
2. Inserting collocative structure initial or terminal indicators
3. Collocative structure segmentation
4. Translation of collocative words
5. Reordering within the collocative structure, and masking of collocative structure
6. Reordering the collocative structure on a syntactical level.

4.3.2.1 Ambiguity Resolving of Collocatives

Words such as 向 (L/I), 来 (L/V/E), 里 (L/S), and 下 (L/V/O), which have tag words besides collocative, are parsed in linguistic pass 1 and their ambiguities are resolved if possible. This process can be illustrated by giving the following rules and their examples:

1. $\$\overline{\text{UX}}$ + LTC/VTA ~ ^{16} /OD + KXX → $\$\overline{\text{UX}}$ + LTC + KXX

这种情况 下 ，

(under this kind of situation,)

下 is a word that can belong to more than one word class: collocative, verb, or O class. In the above rule, with anything except a numeral preceding 下, which is followed by a KXX (phrase or structure indicator), 下 is a collocative, as in the phrase illustrated above. If a numeral precedes 下, it is an OD, as in 一下 (once) and 两下 (twice).

2. NTXXXXXXUX + LTE/VBE ~ ^{16} /EIB0 + KCI → NTXXXXXXUX + LTE + KCI

四十年 来 ，

(for more than forty years,)

来 is a word with more than one class: collocative, verbal, or converbal. In the above rule, when it is preceded by a nominal of time numeral and followed by a KCI phrase indicator, it is a collocative, such as in 三个月来,(for three months,) and in 二十年来,(for twenty years,).

4.3.2.2 Inserting Collocative Structure Initial or Terminal Indicators

The insertion of collocative structure initial (KLI) or terminal (KLT) indicators is processed in linguistic pass 1. A minor indicator (KMX) is inserted after a collocative initial and before a collocative terminal. This is to segment the elements within the collocative structure for reordering purposes in a later process. The following rules are utilized in this set of operations:

1. LIX → KLI + LIX + KMX
2. LTX → KMX + LTX + KLT

4.3.2.3 Collocative Structure Segmentation

Further processing of insertion of collocative structure initial or terminal indicators is necessary when the collocative structure contains only a collocative initial word or only a collocative terminal word. In this phase of operation, structure patterns are matched to determine insertion of indicators. In the course of the analysis some patterns have been found, but they do not include all possible patterns.

1. $\mathrm{KII} + \mathrm{IXXXX}_1 + \mathrm{KMX} + N \sim {}^{10}_{2} + \mathrm{KMX} + \mathrm{LTF}_3 + \mathrm{KLT} \rightarrow \mathrm{KLI} + \mathrm{LTF}_3 + \mathrm{KMX}$
   对 劳动人民 説来
   (to workers)

   $+ N \sim {}^{10}_{2} + \mathrm{KLT}$

2. $\mathrm{KCT} + \mathrm{KLI} + \mathrm{LIB} + \mathrm{KMX} + N \sim {}^{10} + V \sim {}^{16}$
   , 向 美国共产党和美国进步力量 进行
   (proceeds cruel persecution toward the U.S. Communist

   $+ \$\left[\frac{\overline{\mathrm{HM}}}{\mathrm{KXX}}\right] + \text{UP TO NEXT} + \begin{bmatrix}\mathrm{KCI} \\ \mathrm{KPT}\end{bmatrix} \rightarrow \mathrm{KCT} + \mathrm{KLI} + \mathrm{LIB} + \mathrm{KMX} + N \sim {}^{10}$
   残暴 的迫害 ,
   Party and progressive force of the United States)

   $+ \mathrm{KLT} + \mathrm{KVI} + V \sim {}^{16} + \mathrm{KVT} + \$\left[\frac{\overline{\mathrm{HM}}}{\mathrm{KXX}}\right] + \text{UP TO NEXT} + \begin{bmatrix}\mathrm{KCI} \\ \mathrm{KPT}\end{bmatrix}$

The above example illustrates that search must be made beyond the collocative structure itself to determine the collocative structure's terminal point.

3. KLI + LIX + KMX + $\$\left[\frac{\overline{\mathrm{LTX}}}{\mathrm{LIX}}\right]$ + UP TO NEXT + NT3S0N00I0 + KCI

在 积极挑起和扩大局部性的侵略战争 的同时 ，

(at/in the same time of stirring up positively and expanding partial agressive war,)

→ KLI + LIX + KMX + $\$\left[\frac{\overline{\mathrm{LTX}}}{\mathrm{LIX}}\right]$ + UP TO NEXT + NT3S0N00I0 + KLT + KCI

4. KLI + LIX + $\$\left[\frac{\overline{\mathrm{LIX}}}{\mathrm{LTX}}\right]$ + UP TO NEXT + KMX + LTX + KLT

在 同反动統治集团及其代理人的斗争 中

(in struggle with reactionary dominating group and its agent)

→ KLI + LIX + LTX + KMX + $\$\left[\frac{\overline{\mathrm{LIX}}}{\mathrm{LTX}}\right]$ + UP TO NEXT + KLT

The above example illustrates a collocative structure where both collocative initial and terminal words are present in the structure. The terminal word is reordered to follow the initial word so that the translation lookup for the collocative words can be accomplished.

4.3.2.4 Translation of Collocative Words

In this pass, a table has been prepared so that the proper translation for the collocative initial and terminal words can be looked up. Each collocative initial word has two translations: LI1 and LI2 in dictionary III. Each collocative terminal word has four translations: LT1, LT2, nominal, or adjectival-adverbial. To look for the proper translation of LIC + LTD (从 + 看来), the table gives a rule: LIC + LTD → LI1 + LT1. In dictionary III, the LI1 entry for从 has no translation, and the LT1 entry for看来 is translated as "according to." The proper translation for 从 + 看来 is therefore "according to." To look for proper translation of LIC alone, the table gives a rule: LIC → LI2. In dictionary III, the LI2 entry for从 is translated as "from."

4.3.2.5 Reordering Within the Collocative Structure; Masking of Collocative Structure

Within the collocative structure there may be elements such as nouns, noun phrases, relative clauses, or clauses. These must be reordered and grammatic subtags must be inserted to give the proper translation.

After the elements are reordered within the collocative structure, they are masked and the entire collocative structure is ready for syntactical reordering.

To illustrate this process of reordering and masking, the examples given in Section 4.3.2.3 are utilized. In example 1, the structure indicators have been inserted, and the collocative terminal

(LTF) has been reordered to follow the structure initial indicator (KLI). The collocative structure appeared as follows:

1. $\mathrm{KLI + LTF + KMX + NXXXXXXXXX + KLT - KLI + LTF + \underline{KMX + NXXXXXXXOXV} + KLT}$

对 劳动人民 説来　　(to　　workers)

Since the element within this structure is a nominal, it requires no reordering. It is masked for subsequent operations.

In example 2, a rule of similar nature applies, since the element within the structure is a nominal phrase that has been reordered in the nominal phrase operations:

2. $\mathrm{KLI + LIB + KMX + NXXXXXXXXX + KLT - KLI + LIB + \underline{KMX + NXXXXXXXOXX} + KLT}$

向 美国共产党和美国进步力量　　(toward　the U. S. Communist Party
and progressive force of
the United States)

In example 3, the results of the segmentation pass give the following collocative structure:

3. $\mathrm{KLI + LXX_1 + (LXX) + KMX + (KVI) + V \sim {}^{16}_{2} + (KVT) + JC_3 + (KVI) + V \sim {}^{16}_{4} + (KVT)}$

在　　积极挑起　　和　　扩大

$\mathrm{+ N \sim {}^{10}_{5} + NT3S0N00I0_6 + KLT -}$

局部性的侵略戦争　　的同时

The reordering of the elements within the structure is as follows:

4. $\mathrm{KLI + LXX_1 + \underline{(LXX) + KMX + NT3S0N00I0_6 + of + VXXXXXXXXXXXXPXXX_2 + JC_3}}$

(at/in　　the same time　of　stirring up positively　　and

$\mathrm{+ \underline{VXXXXXXXXXXXXPXXX_4 + NXXXXXXOXX_5} + KLT}$

expanding　　partial agressive war)

In example 4 of Section 4.3.2.3 (example 5 below), the element within the structure is a nominal phrase that has been reordered in the nominal phrase operations. Therefore, only masking is required in this pass.

$$5.\ \text{KLI} + \text{LIX} + \text{LTX} + \text{KMX} + \$\left[\frac{\overline{\text{LIX}}}{\text{LTX}}\right] + \quad \text{UP TO NEXT} \quad + \text{KLT} - \text{KLI} + \text{LIX} + \underline{\text{LTX}}$$

在　同反动統治集团及其代理人的斗争　中　(in

$$\underline{+ \text{KMX} + \$\left[\frac{\overline{\text{LIX}}}{\text{LTX}}\right] + \text{UP TO NEXT}} + \text{KLT}$$

struggle with reactionary
dominating group and
its agent)

The noun phrase collocative, such as 在中国的人民 (people in China), is also examined on a syntactical level and reordered in this phase.

4.3.2.6 Reordering of Collocative Structure on a Syntactical Level

On a syntactical level no reordering is necessary if the collocative structure is: (1) at the beginning of a sentence and followed by a comma, (2) preceded by a comma and followed by a period, or (3) preceded and followed by a comma. The collocative structure is reordered to follow the verb or the verb and its object. In all cases, the entire collocative structure is masked in this phase, and no further analysis on this type of structure is made for future passes.

The general operations on collocative structure have been defined in this system. However, to cover all the possibilities of the structures within the collocatives as well as the syntactical qualities of the collocative structure, concordances and further research should be made in this area.

4.4 CONNOMINAL PHRASES

Connominal phrases are adverbial phrases that are introduced by connominals but do not have adverbial phrase ending words. These phrases can also be introduced by connominals and ended by collocative terminals. The connominals used to introduce such phrases have five tags. Tag 1 indicates the major class; tag 2 is an alphabetical designation to distinguish similar groups of connominals; tag 3 indicates whether the connominals need special lookup for a correct English translation of those connominals that have more than one English equivalent; tag 4 indicates the position of the connominals in relation to the word concerned, i.e., it indicates whether they are preverbal, postverbal, preadjectival, etc.; and tag 5 is used to indicate the major structure of the phrase following the connominals involved, i.e., the connominals may be followed by a noun, a noun phrase, a verb, a verb phrase, etc.

The problems involved in connominal phrase analysis are as follows:

1. Ambiguity resolution of connominals that have more than one tag word, e.g., 同 can be a connominal as well as a conjunction (I/J)
2. Determining the beginning and end of a connominal phrase and the insertion of connominal phrase initial and terminal indicators (KII, KIT)
3. Syntactical reordering of connominal phrases and syntactical operations to determine the proper translation of the connominals.

In this year's study, some aspects of these problems have been solved. However, for expansion of tags and rules, further research in these areas should be done.

4.4.1 Ambiguity Resolving of Connominals

In linguistic pass 1, an attempt is made to resolve ambiguities concerning connominals that have more than one tag word. This is done by looking up a table of linguistic rules for a certain sequence of words and resolving these ambiguities by deleting the tag word that is inappropriate to the sequence. For example:

$$\bullet\ \$\left[\frac{\overline{\text{RX}}}{\text{UX}}\right] + \underset{\text{把}}{\text{SU/IFXXX}} \rightarrow \$\left[\frac{\overline{\text{RX}}}{\text{UX}}\right] + \text{IFXXX}$$

In this example, anything that precedes SU/IFXXX (把) except arithmates and numerals resolves the ambiguity to IFXXX. Therefore, in the sentence 他把这件事忘了 (he has forgotten this matter), SU/IFXXX is changed to IFXXX because 他 (he) is neither arithmate nor numeral. The IAAAN/JC (同,跟,和 ,与) can be resolved at this level through a series of sequential rules.

1. $\text{IAAAN/JC} + \text{V}^{*} \sim {}^{16} + \$ + \text{UP TO NEXT} + \text{V} \sim \overset{11}{0} \sim {}^{16} \rightarrow \text{JC} + \text{V} \sim {}^{16} + \$ + \text{UP TO NEXT}$

   和　支持　世界上　被压迫 (and support ....

   $+ \text{V} \sim \overset{11}{0} \sim {}^{16}$

   ....oppress)

---

* The V's in these six rules include A/V and V/N.

2. $\text{IAAAN/JC} + V \sim^{16} + \$ + \text{UP TO NEXT} + V \sim \$\overset{11}{\overline{0}} \sim^{16} \rightarrow \text{IAAAN} + V \sim^{16} + \$$

同　反抗　共产主义的人　斗争　(with　resist

$+ \text{ UP TO NEXT } + V \sim \$\overset{11}{\overline{0}} \sim^{16}$

Communism person　struggle)

3. $\text{IAAAN/JC} + \$\overline{V} \sim^{16}$ (immediately following I/J) $+ \$ + \text{UP TO NEXT} + V \sim \overset{11}{0} \sim^{16} \rightarrow \text{JC}$

和　拉丁美洲　推行　(and

$+ \$\overline{V} \sim^{16}$ (immediately following I/J) $+ \$ + \text{UP TO NEXT} + V \sim \overset{11}{0} \sim^{16}$

Latin America　promote)

4. $\text{IAAAN/JC} + \$\overline{V} \sim^{16}$ (immediately following I/J) $+ \$ + \text{UP TO NEXT} + V \sim \$\overset{11}{\overline{0}} \sim^{16}$

同　反动集团及其代理人的　斗争

$\rightarrow \text{IAAAN} + \$\overline{V} \sim^{16}$ (immediately following I/J) $+ \$ + \text{UP TO NEXT} + V \sim \$\overset{11}{\overline{0}} \sim^{16}$

(with　reactionary dominating group　and its agent　struggle)

5. $\text{IAAAN/JC} + V \sim^{16} + \$\overline{V} \sim^{16} + \text{UP TO NEXT} + \text{KXX} \rightarrow \text{JC} + V \sim^{16}$

和　扩大　局部性的侵略戦争的同时　，

(and　expand

$+ \$\overline{V} \sim^{16} + \text{UP TO NEXT} + \text{KXX}$

partial agressive war　the same time　,)　.

6. $\text{IAAAN/JC} + \$\overline{V} \sim^{16} + \text{UP TO NEXT} + \text{KXX} \rightarrow \text{JC} + \$\overline{V} \sim^{16} + \text{UP TO NEXT}$

和　美国人民大众　.

(and　U.S. people masses

$+ \text{ KXX}$

.)

The six rules indicated above initiate a search for verbs or phrase indicators that resolve the ambiguity of IAAAN/JC. For the first two rules, if there is a verb immediately following

IAAAN/JC, tag 11 of the second verb resolves the ambiguity. For the third and fourth rule, if there is no verb immediately following IAAAN/JC, look for the next series of words to find a verb. If a verb is found, tag 11 of the verb resolves the ambiguity of IAAAN/JC. For the fifth rule, if a verb immediately follows IAAAN/JC and there is no other verb present up to the next structure indicator (KXX), the ambiguity is resolved as conjunction (JC). For the sixth rule, if there is no verb following immediately after IAAAN/JC and up to the next structure indicator (KXX), the ambiguity is resolved as conjunction (JC). More rules will be added to this set of ambiguities.

4.4.2 Determining the Beginning and End of Connominal Phrases and Inserting of Connominal Phrase Initial and Terminal Indicators

In linguistic pass 1, connominal phrase initial indicators (KII) are inserted before a connominal word, and a minor indicator (KMX) is inserted after the connominal word. The minor indicator is necessary, since it aids in the parsing of noun phrases, verb phrases, relative clauses, etc., in subsequent passes. The rule utilized is as follows:

- IXXXX → KII + IXXXX + KMX

根据 (according to)

To find the terminal point of the connominal phrase, this operation is made after ambiguity resolving and after forming of noun phrases, verb phrases, and collocative structures. Some of the situations by which the terminal point of the connominal phrase is found are as follows:

$$1.\ \text{KII} + \text{IFXXX} + \text{KMX} + \text{N} \sim {}^{10} + \text{VTA/N} + \$\left[\begin{array}{c}\bar{\text{V}}\\ \hline \text{HM}\\ \hline \text{V/N}\end{array}\right] + \quad \text{UP TO NEXT} \quad + \left[\begin{array}{c}\text{KCI}\\ \text{KPT}\end{array}\right]$$

把 军费开支 增加 到和平时期的最高峰

$$\rightarrow \text{KII} + \text{IFXXX} + \underline{\text{KMX} + \qquad \text{N} \sim {}^{10} \qquad} + \text{KIT} + \text{KVI} + \text{VTA} \sim {}^{16} + \text{KVT} + \$\left[\begin{array}{c}\bar{\text{V}}\\ \hline \text{HM}\\ \hline \text{V/N}\end{array}\right]$$

$\underbrace{\text{(military expenditures}}_{\text{(IP)*}}$ increase

$$+ \quad \text{UP TO NEXT} \quad + \left[\begin{array}{c}\text{KCI}\\ \text{KPT}\end{array}\right]$$

to highest point of peace period)

---

* The word IP indicates a connominal phrase.

$$2.\ \text{KII} + \text{IXXXX} + \text{KMX} + \qquad N\sim^{10} \qquad + \begin{bmatrix}\text{KPT}\\ \text{KCI}\\ \text{KWI}\\ \text{KVI}\end{bmatrix} \rightarrow \text{KII} + \text{IXXXX} + \underline{\text{KMX}}$$

把　垄斷集团的經济困难　轉嫁....

$$\underline{+ \qquad\qquad N\sim^{10} \qquad\qquad} + \text{KIT} + \begin{bmatrix}\text{KPT}\\ \text{KCI}\\ \text{KWI}\\ \text{KVI}\end{bmatrix}$$

$$\underbrace{\text{(economic difficulty of monopolistic group}}_{\text{(IP)}} \qquad \text{shift....)}$$

$$3.\ \text{KII} + \text{IKXXV} + \text{KMX} + V\sim^{16} + N\sim^{10} + \text{JC} + V\sim^{16} + N\sim^{10} + \begin{bmatrix}\text{KXX}\\ V\end{bmatrix} \rightarrow \text{KII}$$

为　控制　世界　和　发动　世界大戦　开拓....

$$+ \text{IKXXV} + \underline{\text{KMX} + V\sim \overset{13}{P}\sim^{16} + N\sim \overset{8}{O}\sim^{10} + \text{JC} + V\sim \overset{13}{P}\sim^{16} + N\sim \overset{8}{O}\sim^{10}} + \text{KIT} + \begin{bmatrix}\text{KXX}\\ V\end{bmatrix}$$

$$\underbrace{\text{(for} \qquad \text{controlling} \quad \text{world} \quad \text{and} \quad \text{initiating world war}}_{\text{(IP)}} \qquad \text{open....)}$$

After the connominal phrase terminal point is found and the connominal phrase terminal indicator (KIT) is inserted, words may be reordered and tags may be changed within the connominal phrase (see example 3 above). The tag words from KMX to the last tag word of the connominal phrase are masked, and KII + IXXXX + KLT is left for reordering on the syntactical level and for I translation lookup.

4.4.3 Reordering of Connominal Phrases on a Syntactical Level

On a syntactical level, no reordering is necessary if the connominal phrase is: (1) at the beginning of a sentence and followed by a comma, (2) preceded by a comma and followed by a period, or (3) preceded and followed by a comma. Generally, if it precedes the verb in the processing sentence, the connominal phrase is reordered to follow the verb or the verb and its object. If a connominal phrase begins with the connominals IK (为), IN (成), or IO (作, 作为, 做) and the phrase follows the verb in the processing sentence, the verb subtag 8 or 10 is examined for the correct rules for translation and reordering. In the same pass, the proper translation of connominals will be looked up by reference to the verb connominal subtags. For regular connominals, i.e., connominals whose tag 3 is R (regular translation), no translation lookup is necessary. For irregular connominals, i.e., connominals whose tag 3 is I (irregular translation), the verb connominal subtags 6, 7, 8, and 11 are referred to so that the proper translation may be determined.

Further study will be made in this area concerning other discontinuous structures in relation to connominal phrases, expansion of connominal subtags, and elements within the connominal phrases.

4.5 NOUN PHRASE-RELATIVE CLAUSES

The noun phrase-relative clause,* in the Chinese sense, is a noun or noun phrase modified by a subordinate clause, which consists of a verb with or without subject and object. The verb may take the participle, present, or passive forms in English, depending on the pattern and content of the relative clause in Chinese. The entire relative clause becomes a noun phrase in relation to the sentence structure. The word order in the Chinese pattern is different from the English, since the head word is the last word in the pattern and is usually preceded by HM (的). In Chinese, the subordinate clause precedes the head word and is connected to the head word by HM. In English, the head word precedes the subordinate clause, and the subordinate clause is introduced by expressions such as "of," "that," and "by which."

The problems involved in relative clause analysis are twofold. First, patterns of relative clauses must be found and the operations must be determined to reorder the Chinese words and to alter their subtags to indicate the varied forms these words will take in English. Second, the beginning and end of the relative clause in the sentence structure must be determined and indicated by structure initial and terminal indicators. Within this phase of operation, the head word must be singled out for subsequent verb linking passes, since the relative clause behaves like a nominal in the syntactic structure.

4.5.1 Noun Phrase-Relative Clause Patterns

The first problem is to determine the various relative clause patterns that can be found in the sentence structure. In this study, patterns have been found whose translation and word reorder vary accordingly to the quality of both the head word and the verb. These patterns are therefore further subdivided for proper translation according to the quality of the head word and the verb.

The following relative clause patterns are listed, and they are subdivided according to the quality of the verb and the head word. Word reordering, subtag altering, and proper English connecting word adding is indicated.

---

* The terms "noun phrase-relative clause" and "relative clause" are used interchangeably in this report.

1. **VXA Verbs:**

- VXA + N_1 + HM + N_2 (D)* → N_2 + of + VXAXXXXXXXXXPXXX + N_1

反抗共产党的斗争 (struggle of resisting Communist Party)

- VXA + N_1 + HM + NXX$\begin{bmatrix}S\\P\end{bmatrix}_2$ (H, O, I, E, C, B, P) → NXX$\begin{bmatrix}S\\P\end{bmatrix}_2$ + that + VXAXXXXXXXXXX$\begin{bmatrix}S\\P\end{bmatrix}$XX + N_1

反抗共产党的人民 (people that resist Communist Party)

- NXX$\begin{bmatrix}S\\P\end{bmatrix}_1$ + VXA + N_2 + HM + N_3 → N_3 + by which + NXX$\begin{bmatrix}S\\P\end{bmatrix}_1$ + VXAXXXXXXXXXX$\begin{bmatrix}S\\P\end{bmatrix}$XX + N_2

政府管理国家的政策 (policy by which Government manages country)

- ER00 (所) + VXA + HM + NXX$\begin{bmatrix}S\\P\end{bmatrix}$ → NXX$\begin{bmatrix}S\\P\end{bmatrix}$ + that + VXAXXXXXXXXXX$\begin{bmatrix}S\\P\end{bmatrix}$XP

所減低的工資 (wages that are decreased)

- NXX$\begin{bmatrix}S\\P\end{bmatrix}_1$ + ER00 (所) + VXA + HM + N_2 → N_2 + that + NXX$\begin{bmatrix}S\\P\end{bmatrix}_1$ + VXAXXXXXXXXXX$\begin{bmatrix}S\\P\end{bmatrix}$XX

公司所減低的工資 (wages that company decreases)

- NXX$\begin{bmatrix}S\\P\end{bmatrix}_1$ + VXA + HM + N_2 → N_1 + that + NXX$\begin{bmatrix}S\\P\end{bmatrix}_2$ + VXAXXXXXXXXXX$\begin{bmatrix}S\\P\end{bmatrix}$XX

公司減低的工資 (wages that company decreases)

---

* The symbols within the parentheses indicate the second tag of the nominal.

- EAI0 (被) + VXA + HM + NXX$\begin{bmatrix}S\\P\end{bmatrix}$ – NXX$\begin{bmatrix}S\\P\end{bmatrix}$ + that + VXAXXXXXXXXXXX$\begin{bmatrix}S\\P\end{bmatrix}$XP

被消灭的房子 (house that is destroyed)

- NXX$\begin{bmatrix}S\\P\end{bmatrix}_1$ + EAI0 (被) + VXA + HM + N_2 – N_2 + that + NXX$\begin{bmatrix}S\\P\end{bmatrix}_1$

他被殺的原因 (reason that he

+ VXAXXXXXXXXXXX$\begin{bmatrix}S\\P\end{bmatrix}$XP

is killed)

- IERBN (給) + NXX$\begin{bmatrix}S\\P\end{bmatrix}_1$ + VXA + HM + N_2 – N_2 + that + NXX$\begin{bmatrix}S\\P\end{bmatrix}_1$

給水災消灭的人 (people that flood

+ VXAXXXXXXXXXXX$\begin{bmatrix}S\\P\end{bmatrix}$XX

destroys)

- VXA + HM + NXX$\begin{bmatrix}S\\P\end{bmatrix}$ (I, O, C, B, P) – NXX$\begin{bmatrix}S\\P\end{bmatrix}$ + that + VXAXXXXXXXXXXX$\begin{bmatrix}S\\P\end{bmatrix}$XP:

洗的衣服 (clothes that are washed); 改組的共产党 (Communist Party that is reorganized); 开发的非洲 (Africa that is developed); 養的狗 (dog that is reared)

- VXA_1 + HM + ND_2 – ND_2 + of + $NXXXXXXXXXV_1$

反抗的思想 (thought of resistance)

- VXA + HM + NXX$\begin{bmatrix}S\\P\end{bmatrix}$ (H, E) – NXX$\begin{bmatrix}S\\P\end{bmatrix}$ + that + VXAXXXXXXXXXXX$\begin{bmatrix}S\\P\end{bmatrix}$XX

洗的人 (people that wash).

2. VI0 Verbs:

- $\text{VI0} + \text{HM} + \text{NXX}\begin{bmatrix}S\\P\end{bmatrix}(\text{H, E, O, B, I, C, P}) \rightarrow \text{NXX}\begin{bmatrix}S\\P\end{bmatrix} + \text{that} + \text{VI0XXXXXXXXXXX}\begin{bmatrix}S\\P\end{bmatrix}\text{XX}$:

  跑的人 (people that run); 独立的国家 (country that is independent); 跑的馬 (horses that run); 生长的樹木 (trees that grow)

- $\text{VI0}_1 + \text{HM} + \text{ND}_2 \rightarrow \text{N}_2 + \text{of} + \text{NXXXXXXXXXXV}_1$

  备戦的理論 (theory of preparing war).

3. VXBI Verbs:

- $\text{VXBI}_1 + \text{N}_2 + \text{V}\begin{bmatrix}I\\B\end{bmatrix}_3 + \text{HM}_4 + \text{NXX}\begin{bmatrix}S\\P\end{bmatrix}_5 \rightarrow \text{NXX}\begin{bmatrix}S\\P\end{bmatrix}_5 + \text{that} + \text{VXBIXXXXXXXXXX}\begin{bmatrix}S\\P\end{bmatrix}\text{XX}_1 + \text{NXXXXXXOXX}_2 + \text{VXXXXXXXXXXXXIXXX}_3$: 使他去的原因 (reason that causes him to go); 准他走的人 (people that permit him to go)

- $\text{VXBI}_1 + \text{N}_2 + \text{AJ}_3 + \text{HM}_4 + \text{NXX}\begin{bmatrix}S\\P\end{bmatrix}_5 \rightarrow \text{NXX}\begin{bmatrix}S\\P\end{bmatrix}_5 + \text{that} + \text{VXBIXXXXXXXXXX}\begin{bmatrix}S\\P\end{bmatrix}\text{XX}_1 + \text{NXXXXXXOXX}_2 + \text{to be} + \text{AJ}_3$: 使他高兴的事情 (matters that cause him to be happy)

- $\text{VXBI}_1 + \text{N}_2 + \text{VXA}_3 + \text{HM}_4 + \text{NXX}\begin{bmatrix}S\\P\end{bmatrix}_5 \rightarrow \text{NXX}\begin{bmatrix}S\\P\end{bmatrix}_5 + \text{that} + \text{VXBIXXXXXXXXXX}\begin{bmatrix}S\\P\end{bmatrix}\text{XX}_1 + \text{NXXXXXXOXX}_2 + \text{VXXXXXXXXXXXXIXXX}_3$: 帮助美国进攻的决心 (determination that helps U.S. to attack); 請先生解釋的学生 (students that request teacher to explain)

- $\text{N}_1 + \text{VXBI}_2 + \text{N}_3 + \text{VI0}_4 + \text{HM}_5 + \text{N}_6 \rightarrow \text{N}_6 + \text{of} + \text{NXXXXXXPXX}_1 + \text{VXBIXXXXXXXXPXXX}_2 + \text{NXXXXXXOXX}_3 + \text{VXXXXXXXXXXXXIXXX}_4$: 我们使他去的原因 (reason of our causing him to go); 美国使日本投降的工具 (tools of the U.S.'s causing Japan to surrender)

- $N_1 + VXBI_2 + N_3 + AJ_4 + HM_5 + N_6 - N_6 + \text{of} + NXXXXXXXPXX_1$

  $+ VXBIXXXXXXXXXPXXX_2 + NXXXXXXXOXX_3 + \text{to be} + AJ_4$: 人民使社会繁荣的計画(plan of people's causing society to be prosperous)

- $N_1 + VXBI_2 + N_3 + VXA_4 + HM_5 + N_6 - N_6 + \text{of} + NXXXXXXXPXX_1$

  $+ VXBIXXXXXXXXXPXXX_2 + NXXXXXXXOXX_3 + VXXXXXXXXXXXXXIXXX_4$: 我们帮助美国进攻的决心(determination of our helping U.S. to attack); 先生鼓勵学生发问的方法(method of teacher's encouraging students to ask questions).

4. VTBW Verbs:

- $VTBW + N_1 + HM + NXX\begin{bmatrix}S\\P\end{bmatrix}_2 - NXX\begin{bmatrix}S\\P\end{bmatrix}_2 + \text{that} + VTBWXXXXXXXXXXXXP + \text{to}$

  $+ NXXXXXXXOXX_1$: 給他的信 (letter that is given to him); 付他的钱 (money that is paid to him)

- $NXX\begin{bmatrix}S\\P\end{bmatrix}_1 + VTBW_2 + N_3 + HM_4 + N_5 - N_5 + \text{that} + NXX\begin{bmatrix}S\\P\end{bmatrix}XXXSXX_1$

  $+ VTBWXXXXXXXXXX\begin{bmatrix}S\\P\end{bmatrix}XX_2 + NXXXXXXXOXX$: 我給他的信 (letter that I give him); 我付他的钱 (money that I pay him).

The above patterns would be relevant only if noun phrase-relative clause initial and terminal indicators had been indicated in the processing of linguistic rules.

4.5.2 Noun Phrase-Relative Clause Operations

The noun phrase-relative clause operations consist of: (1) determining the beginning and end of the noun phrase-relative clauses and designating initial and terminal indicators for these structures, and (2) reordering within the relative clause to give the proper translation and masking all words except the head word for verb linking. Environments of the relative clause must be examined to determine the initial and terminal of the relative clause. It is possible that the relative clause may contain collocative structures and/or connominal phrases. This phase of work needs more detailed analysis for the future. As indicated in Section 4.5.1, there are many noun phrase-relative clause patterns in which words are reordered to derive the proper translation, depending on the content of the tag words. Future studies may produce more patterns than are shown in this report.

The following examples demonstrate the process of locating the initial and terminal points of the relative clause and the reordering and masking of these structures.

$$1.\ \begin{bmatrix}\text{KCT}\\ \text{KPI}\end{bmatrix} + \quad \text{N} \quad + \text{VT}\begin{bmatrix}\text{F}\\ \text{A}\end{bmatrix} + \quad \text{N} \quad + \text{HM} + \quad \text{N} \quad + \text{KCI} + \text{PC}$$

，　各阶层的人民 反对 军国主义化　的　广泛的斗争　，

$$+ \text{KCT} + \quad \text{KVI} + \text{VTA} \quad \rightarrow \begin{bmatrix}\text{KCT}\\ \text{KPI}\end{bmatrix} + \text{KRI} + \text{N}_1 + \text{VT}\begin{bmatrix}\text{F}\\ \text{A}\end{bmatrix}_2 + \text{N}_3 + \text{HM}_4 + \text{N}_5 + \text{KRT}$$

正在蓬勃地发展

$$+ \text{KCI} + \text{PC} + \text{KCT} + \text{KVI} + \text{VTA}^{*} \rightarrow \begin{bmatrix}\text{KCT}\\ \text{KPI}\end{bmatrix} + \text{KRI} + \text{NXXXXXRXXX}_4 + \underline{\text{by which}}$$

(extensive struggle　　by which

$$+ \underline{\text{NX}^{3}\,{}^{4}\text{XXXXXX}_1 \quad + \text{VT}\begin{bmatrix}\text{F}\\ \text{A}\end{bmatrix}\text{XXXXXXXXXX}^{13\ 14\dagger}\ \text{XX}_2}$$

people of various strata　　　　oppose

$$+ \underline{\text{NXXXXXXOXX}_3 + \text{KRT} + \text{KCI} + \text{PC} + \text{KCT} + \text{KVI} + \text{VTA}^{\ddagger}}$$

militarization)

The above example first gives the syntactical pattern as the argument. The first function inserts the noun phrase-relative clause initial and terminal indicators. The second function reorders the elements within the noun phrase-relative clause and adds English words for proper translation. Subtags may also be changed, as illustrated in this example. In this structure, the person and number tags (tags 3 and 4) of N_1 are duplicated for the person and number tags (tags 13 and 14) of the verb. The elements within the structure are then masked, and only the head word is left for subsequent passes.

---

[*] This function inserts noun phrase-relative clause initial and terminal indicators (KRI, KRT).

[†] Tags 3 and 4 of the nominal are duplicated for tags 13 and 14 of the verbal.

[‡] This function reorders tag words, adds English words, alters subtags, and masks elements within the noun phrase-relative clause.

2. $\begin{bmatrix}\text{KCT}\\\text{KPI}\end{bmatrix}$ + N + KWI + VTD + KWT + HM

肯尼迪之流 所喧嚷 的

+ N + $\begin{bmatrix}\text{KCI}\\\text{KPT}\end{bmatrix}$

什么"尊重人权"、"人人得到自由和平等" ,

$\rightarrow \begin{matrix}\text{KCT}\\\text{KPI}\end{matrix} + \text{KRI} + \text{N}_1 + \text{KWI} + \text{VTD}_2 + \text{KWT} + \text{HM}_3 + \text{N}_4 + \text{KRT} + \begin{bmatrix}\text{KCI}\\\text{KPT}\end{bmatrix} \rightarrow \begin{bmatrix}\text{KCT}\\\text{KPI}\end{bmatrix} + \text{KRI}$

$+ \text{NXXXXXRXXX}_4 + \underline{\text{which} + \text{NX}^{3}\,{}^{4}\text{XXXSXX}_1}$

("respect of human rights," every- which Kennedy's category

one obtains liberty and equality

$+ \underline{\text{VXXXXXXXXXXX}^{13}\ {}^{14}\text{XA}_2} + \text{KRT} + \begin{bmatrix}\text{KCI}\\\text{KPT}\end{bmatrix}$

cries)

With the above syntactical pattern as the argument, the first function and the second function perform the operations as explained in example 1. However, since the verb is VTD, the English word inserted is "which" instead of "by which," and N_1 takes the subjective case and the verb is in the active voice.

3. $\begin{bmatrix}\text{KCT}\\\text{KPI}\end{bmatrix}$ + N + VTA + KLI + LXX + KLT + HM

肯尼迪政府 加 在美国共产党头上 的

$+ \text{N} + \text{KCI} \rightarrow \begin{bmatrix}\text{KCT}\\\text{KPI}\end{bmatrix} + \text{KRI} + \text{N}_1 + \text{VTA}_2 + \text{KLI} + \text{LXX}_3 + \text{KLT} + \text{HM}_4 + \text{N}_5 + \text{KRT}$

罪状 ,

$+ \text{KCI} \rightarrow \begin{bmatrix}\text{KCT}\\\text{KPI}\end{bmatrix} + \text{KRI} + \text{NXXXXXRXXX}_5 + \underline{\text{that} + \text{NX}^{3}\,{}^{4}\text{XXXSXX}_1}$

(crimes that Kennedy's government

$+ \underline{\text{VTAXXXXXXXXXX}^{13}\ {}^{14}\text{XA}_2 + \text{KLI} + \text{LXX}_3} + \text{KLT} + \text{KRT} + \text{KCI}$

adds on the U.S. Com-

munist Party)

The above example illustrates the occurrence of a collocative structure within the noun phrase-relative clause. The collocative structure is reordered to follow the verb. The English word "that" is inserted, since the relative clause pattern shows the proper form of translation for this structure.

4. $\mathrm{VS1} + \mathrm{KVT} + \begin{bmatrix} \mathrm{VTA} \\ \mathrm{VTA} + \mathrm{JC} + \mathrm{VTX} \end{bmatrix} + \quad \mathrm{N}$

是　同情　和　支持　世界上被压迫人民和被压迫

$+ \ \mathrm{HM}$

民族解放斗争　的

$+ \quad \mathrm{N} \quad + \begin{bmatrix} \mathrm{KCI} \\ \mathrm{KPT} \end{bmatrix} \rightarrow \mathrm{VS1} + \mathrm{KVT} + \mathrm{KRI} + \begin{bmatrix} \mathrm{VTA} \\ \mathrm{VTA} + \mathrm{JC} + \mathrm{VTX} \end{bmatrix}_1 + \mathrm{N}_2 + \mathrm{HM}_3 + \mathrm{N}_4$

国际主义者　,

$+ \ \mathrm{KRT} + \begin{bmatrix} \mathrm{KCI} \\ \mathrm{KPT} \end{bmatrix} \rightarrow \mathrm{VS1} + \mathrm{KVT} + \mathrm{KRI} + \mathrm{NX}^3 \ {}^4\mathrm{XXXXXX}_4 + \mathrm{that} + \underline{\begin{bmatrix} \mathrm{VTA} \\ \mathrm{VTAXXXXXXXXXX}^{13} \ {}^{14}\mathrm{XX} \end{bmatrix}}$

(internationalists that sympathize with

$\underline{+ \ \mathrm{JC} + \mathrm{VTXXXXXXXXXXX}^{13} \ {}^{14}\mathrm{XX}\,] + \mathrm{NXXXXXXXXXX}_2} + \mathrm{KRT} + \begin{bmatrix} \mathrm{KCI} \\ \mathrm{KPT} \end{bmatrix}$

and support oppressed people
and oppressed
national liberation
struggle on world ,)

The above example illustrates a specific pattern where the verb 是 (to be) is present. Reordering, insertion, and masking operations were used for this processing.

These are only a few of the noun phrase-relative clause examples and the operations that have been illustrated. The difficulties involved in analyzing this type of structure are in the determination of initial and terminal points. More sentence patterns should be analyzed for a complete processing of this type of structure.

4.6 VERB LINKING

After the processing of major phrases and attributive structures, verbs are linked to the subject. At this time, all attributive structure initial and terminal indicators (KXX) are masked, with the exception of KCI, KCT, KPI, and KPT. Only major punctuation initial and terminal indicators and head words of major phrases remain in the processing sentence. Verb linking sentence patterns are then processed in the form of rules to derive the proper person and number for the verb and to insert subject or connecting words between clauses.

Table 4-1 contains examples of some of the rules we have found in the course of our analyses for subject and verb linking.

4.7 ILLUSTRATIONS OF LINGUISTIC PROCESSING ON A SENTENCE LEVEL

Tables 4-2, 4-3, 4-4, and 4-5 are used to show the step by step linguistic processing utilizing the techniques resulting from the present study. In these examples, sentence 1 (Tables 4-2 and 4-3) is a simple sentence and sentence 2 (Tables 4-4 and 4-5) is a complex sentence. Each sentence is first presented in art form, giving the following information:

1. Original Chinese text
2. General translation for individual Chinese words. The lexicographic entry for each word gives all English forms with different inflections.
3. Chinese grammatic tags for individual Chinese words
4. Indication of changes made through the utilization of linguistic rules
5. Resulting English grammatic tags for individual Chinese words
6. Resulting English translation through the lookup of English grammatic tags
7. Numbers in parentheses in the examples refer to rule numbers in Tables 4-3 and 4-5 and also indicate the precedence of rules.

A list of linguistic rules (Tables 4-3 and 4-5) utilized in the linguistic processing of the sentence follows the sentence illustration. The linguistic pass number is indicated in the left column. The numbers in the right column refer to the precedence of rules.

By using these illustrations it is hoped that the linguistic analysis may be further clarified for a Chinese to English machine translation system as detailed in this report.

Table 4-1 — Verb Linking Sentence Patterns

VTG + , and VTF + ,

1. Look for VS1 in the sentence. If any verb is between VTG + , or VTF + , and VS1 + N + . change the verb to participle form.

$$\bullet\ \mathrm{N}_1 + \begin{bmatrix}\mathrm{VTG}\\ \mathrm{VTF}\end{bmatrix}_2 + ,_3 + (\mathrm{N})_4 + \mathrm{V}_5$$

$$+ (\mathrm{N})_6 + \mathrm{VS1}_7 + \mathrm{N}_8 + ._9 \rightarrow \mathrm{N}_1 + \begin{bmatrix}\overset{\downarrow\ \ *}{\mathrm{VTG}}\\ \mathrm{VTF}\end{bmatrix}_2$$

$$+ ,_3 + (\mathrm{NXXXXXXXPXX})_4$$

$$+ \mathrm{VXXXXXXXXXXXXPXXX}_5 + (\mathrm{NXXXXXXXOXX})_6$$

$$+ \mathrm{VS1XXXXXXXXXX3SXX}_7 + \mathrm{N}_8 + ._9$$

2. If VS1 is not in the sentence:

$$\bullet\ \mathrm{N}_1 + \begin{bmatrix}\mathrm{VTG}\\ \mathrm{VTF}\end{bmatrix}_2 + ,_3 + \mathrm{N}_4 + \ldots.$$

$$\rightarrow \mathrm{N}_1 + \begin{bmatrix}\overset{\downarrow}{\mathrm{VTG}}\\ \mathrm{VTF}\end{bmatrix}_2 + ,_3 + \mathrm{N}_4 + \ldots.$$

$$\bullet\ \mathrm{N}_1 + \begin{bmatrix}\mathrm{VTG}\\ \mathrm{VTF}\end{bmatrix}_2 + ,_3 + \mathrm{V}_4 \ldots.$$

$$\rightarrow \mathrm{N}_1 + \begin{bmatrix}\overset{\downarrow}{\mathrm{VTG}}\\ \mathrm{VTF}\end{bmatrix}_2 + ,_3 + \underline{\mathrm{N}}\ \Big(\text{reproduce}$$

$$\text{the first N preceding } \begin{bmatrix}\mathrm{VTG}\\ \mathrm{VTF}\end{bmatrix}\Big) + \overset{\downarrow}{\mathrm{V}}_4 \ldots.$$

VTA and Other Verbs

$$1.\ \mathrm{N}_1 + \begin{bmatrix}\mathrm{VTBW}\\ \mathrm{VS1}\\ \mathrm{VTA}\end{bmatrix}_2 + \mathrm{N}_3 + ,_4 + \mathrm{V}_5 + \mathrm{N}_6$$

$$\rightarrow \mathrm{N}_1 + \begin{bmatrix}\overset{\downarrow}{\mathrm{VTBW}}\\ \mathrm{VS1}\\ \mathrm{VTA}\end{bmatrix}_2 + \mathrm{N}_3 + ,_4 + \overset{\downarrow}{\mathrm{V}}_5 + \mathrm{N}_6$$

$$2.\ \mathrm{N}_1 + \begin{bmatrix}\mathrm{VTBW}\\ \mathrm{VS1}\\ \mathrm{VTA}\end{bmatrix}_2 + \mathrm{N}_3 + ,_4 + \mathrm{N}_5 + \mathrm{V}_6 \ldots.$$

$$\rightarrow \mathrm{N}_1 + \begin{bmatrix}\overset{\downarrow}{\mathrm{VTBW}}\\ \mathrm{VS1}\\ \mathrm{VTA}\end{bmatrix}_2 + \mathrm{N}_3 + ,_4 + \mathrm{N}_5 + \overset{\downarrow}{\mathrm{V}}_6 \ldots.$$

* An arrow from N to V indicates that the verb derives the number and person tags from the noun.

PRECEEDING PAGE BLANK NOT FILMED

Table 4-2 — Sample Sentence I

*1	I		not	doubt	,	the U.S. Communists	with/and	the United States	people
*2	我	们	毫不	怀疑	,	美国共产党人	和	美国	人民
*3	KPT NM1SPM0000	HN	ENI0	VTF0AP0200000000	PC	NE3PPM0000	IAAAN/JC	NC3SPF0000	NE3P0M0000
*6	NM1PPM0000(4)^{*5}	KVI(7)		VTF0AP02000Y0000(7)	KVT(7) KCI(2)	KCT(2)	JC(3)	AJR00T	
						(13)			
									NE3P0M1000(10)
	NM1PPM0S00(14)^{*7}			VTF0AP02000Y1P00(14)					NE3P0MJ000(13)
									NE3P0MJS00(14)
*4	We			do not doubt	.	the U.S. Communists	and	American	people

possess	boundless	bright	de	future	,	certain	can	obtain	
具有	无限	光明	的	前途	,	一定	能够	取得	
VTA0000000000000	AARJ0T	AJR00T	HM	ND3S0N0000	PC	AARV0T	GN	VTA0002000000000	
	ADRA0T(5)				KCI(2) KCT(2)KVI(8)	ADRV0T$_1$(6)	GN$_2$	VTA0002000000000$_3$	KVT(8)
				ND3S0N1000(11)		GN$_2$	VTA000200000G000$_3$ ADRV0T$_1$		
							VTA000200000G000(8)		
VTA0000000003P00(14)				ND3S0N1O00(14)			VTA000200000GP00(14)		
possess	boundlessly	bright		future	,		can obtain certainly		

struggle/struggle	de	victory		
斗争	的	勝利		
VI00043000100000/NS3S0N0000	HM	ND3S0N0000	PP	
ND3S0N0000$_1$(9)	HM$_2$	ND3S0N0000$_3$	KPT(1)	KPI(1)
ND3S0N0000$_3$	of	ND3S0N0O00$_1$		
ND3S0N2000(12)				
ND3S0N2O00(14)				
victory	of	struggle		

---

*[1] General translation for respective Chinese words. The dictionary entry gives all English forms with different inflect

*[2] Text.

*[3] Grammatic tags for respective Chinese words and changes made through linguistic rules (Table 4-3).

*[4] Resulting translation.

*[5] Number in parentheses refers to rule number in Table 4-3, and also indicates the precedence of rules.

*[6] The underscore indicates that one or more tag words are masked or deleted. See Table 4-3 for details.

*[7] See Section 5 for information concerning how the tags are used to choose correct English forms.

Table 4-3 — Rules Used in Sample Sentence I

Pass	Rules	Rule Number
1A	$\text{PP} - \text{KPT} + \text{PP} + \text{KPI}$	1
	$\text{PC} - \text{KCI} + \text{PC} + \text{KCT}$	2
1D	$\text{IAXXX/JC}_1 + \$_2 + \text{UP TO NEXT}_3 + \text{VXXXXXXXXX1XXXXX} - \text{JC}_1 + \$_2 + \text{UP TO NEXT}_3$ $+ \text{VXXXXXXXXX1XXXXX}_4$	3
1F	$\text{NXXXXXXXXX}_1 + \text{HN}_2 - \text{NXXPXXXXXX}_1$	4
	$\text{AARJXX}_1 + \text{AJRXXX}_2 - \text{ADRAXX}_1 + \text{AJRXXX}_2$	5
	$\text{AAXXXX}_1 + \text{GN}_2 - \text{ADXXXX}_1 + \text{GN}_2$	6
1G	$\text{ENI0}_1 + \text{VXXXXXXXXXXXXXXX}_2 - \text{KVI} + \text{VXXXXXXXXXXYXXXX}_2 + \text{KVT}$	7
	$\text{ADRVXX}_1 + \text{GN}_2 + \text{VXXXXXXXXXXXXXXX}_3 - \text{KVI} + \underline{\text{GN}_2 +}\ \text{VXXXXXXXXXXXGXXX}_3 + \underline{\text{ADRVXX}_1 +}\ \text{KVT}$	8
1H	$\text{KXX}_1 + \text{VXXXXXXXXXXXXXXX/NXXXXXXXXX}_2 + \text{HM}_3 + \text{NDXXXXXXXX}_4 + \text{KXX}_5 - \text{KXX}_1$ $+ \text{NXXXXXXXXX}_2 + \text{HM}_3 + \text{NDXXXXXXXX}_4 + \text{KXX}_5$	9
2A	$\text{NCXXXXXXXX}_1 + \text{NXXXXXXXXX}_2 - \underline{\text{AJROOT}_1} + \text{NXXXXX1XXX}_2$	10
	$\begin{bmatrix}\text{KCT}\\ \text{KPI}\\ \text{KMX}\end{bmatrix}_1 + \$\overline{\text{VXXIXXXXXXXXXXXX}\ \begin{pmatrix}\text{KVT}\\ \text{KWT}\end{pmatrix}_2} + \text{UP TO NEXT} + \text{ADRAXX}_3 + \text{AJXXXX}_4 + \text{HM}_5$ $+ \text{NXXXXXXXXX}_6 - \begin{bmatrix}\text{KCT}\\ \text{KPI}\\ \text{KMX}\end{bmatrix}_1 + \$\overline{\text{VXXIXXXXXXXXXXXX} + \begin{pmatrix}\text{KVT}\\ \text{KWT}\end{pmatrix}_2} + \text{UP TO NEXT} + \underline{\text{ADRAXX}_3}$ $\underline{+ \text{AJXXXX}_4} + \text{NXXXXX1XXX}_6$	11
2B	$\text{KXX}_1 + \text{NXXXXXXXXX}_2 + \text{HM}_3 + \text{NXXXXXXXXX}_4 + \text{KXX}_5 - \text{KXX}_1 + \text{NXXXXX2XXX}_4 + \underline{\text{of}}$ $\underline{+ \text{NXXXXXXOXX}_2} + \text{KXX}_5$	12
	$\text{NXXXXXXXXX}_1 + \text{JC}_2 + \text{NXXXXXXXXX}_3 - \text{NXXXXXXXXX}_1 + \text{JC}_2 + \text{NXXXXXJXXX}_3$	13
5	$\text{KPI}_1 + \text{NX1PXXXXXX}_2 + \text{VTFXXXXXXXXXXXXX}_3 + \text{PC}_4 + \text{NX3PXXXXXX}_5 + \text{VTAXXXXXXXXXXXXX}_6$ $+ \text{NXXXXXXXXX}_7 + \text{PC}_8 + \text{VTAXXXXXXXXXXXXX}_9 + \text{NXXXXXXXXX}_{10} + \text{KPT}_{11} - \text{KPI}_1$ $+ \text{NX1PXXXSXX}_2 + \text{VTFXXXXXXXXX1PXX}_3 + \text{PC}_4 + \text{NX3PXXXSXX}_5 + \text{VTAXXXXXXXXX3PXX}_6$ $+ \text{NXXXXXXOXX}_7 + \text{PC}_8 + \text{VTAXXXXXXXXX3PXX}_9 + \text{NXXXXXXOXX}_{10} + \text{PP}_{11}$	14*

* This is a combination of several rules. For details, see rules in pass 5 of rule file.

PRECEEDING PAGE BLANK NOT FILMED

Tab

*1 in/at the United States in the country , Kennedy g

*2 在 美国 国内 ， 肯尼迪

*3 KPI NC3SPF0000 LTA

KLI(4)*5 LID KMX(4) KMX(4) PC NH3SPM0000 NC

KLI(1) LID LTA KMX NC3SPF0O00 KLT(4)KCI(2) KCT(2) NH3SPM0P00

LI1 LT1(22) KLT1(21) NC

*6 LI1*7 (24) NC

(30)

*4 Inside the United States , Kennedy's gove

increase tax revenue further

增加 税收 、 进一步

PS VTA0B00000000000/ND3S0N0000 ND3S0N0000 PS ADRVOT VBA0

KVI(19) VTA0B00000000000(19) KVT(19) KVI(6) VBA0B0020N000000

KWI(27) KWT(27) KWI(27) VBA0B0020N000000(7)

PS_3 $\mathrm{VTA0B00000000000}_4$ $\mathrm{ND3S0N0000}_5$ PS_6 $\mathrm{VBA0B0020N000000}_7$

$\mathrm{NI3P0N1O00}_2$ PS_3 $\mathrm{VTA0B0000000P000}_4$ $\mathrm{ND3S0N0000}_5$ PS_6 $\mathrm{VBA0B0020N00P000}_7$

wages , increasing tax revenue , lowering

attempt monopolistic group economy difficult/difficulty

企图 把 垄断集团 的 经济 困难

VTE000000000000/ND3S0N0000 SU/IFRRN NO3S0N0000 HM ND3S0N0000 AJR00T/ND3S0N0000

VTE000000000000(10) KII(6) IFRRN(3)KMX(6) AJR00T ND3S0N0000(11) KIT(25) K

VTE00000E0003S00(32) $\mathrm{NO3S0N0000}_1$ HM_2 $\mathrm{ND3S0N1000}_3$(14)

$\mathrm{ND3S0N1000}_3$ of $\mathrm{NO3S0N0O00}_1$

ND3S0N2000(18)

KII_1 IFRRN_2(26) KIT_3(26)

KWI $\mathrm{VTA0B00000000000}_4$ KWT KII_1 IFRRN_2 KIT_3(29)

VTA0B0000000I000(32) NXXXXXXXXX(31)

NXXXXXXOXX(32)

attempts to shift economic difficulty of monopolistic group

---

*1 General translation for respective Chinese words. The dictionary entry gives all English forms with different inflections.

*2 Text.

*3 Grammatic tags for respective Chinese words and changes made through linguistic rules (Table 4-5).

*4 Resulting translation.

Table 4-4 — Sample Sentence II

government adopt decrease actual wages

政府 采取 减低 实际 工资

00 NO3S0N0000 VTA0B20000000000 VTA0CP0200000000 AARV0T NI3P0N0000

KRI(27) KVI(20) KVT(20) AJRV0T

00 KVI(27) KVT(27) KWI(27) KWT(27) NI3P0N1000(9)

NO3S0N1000(12) VTA0CP0200000000$_1$ N13P0N1000$_2$

NO3S0N1S00(32) VTA0B20000003S00(32) ND3S0NR000$_{10}$ of VTA0CP020000P000$_1$

ND3S0NR000(28)

ND3S0NRO00(32)

's government adopts policy of decreasing actual

lower people living standard policy ,

降低 人民 生活水平 的 政策 ，

VBA0B0020N000000 NE3P0M0000 ND3S0N0000 HM ND3S0N0000 PC

ADRV0T KVT(7) NE3P0M0P00 KRT(27) KCI(2) KCT(2)

KWT(26) ND3S0N1000(13)

ND3S0N1000$_8$ HM$_9$ ND3S0N0000$_{10}$

ND3S0N1O00$_8$

further people's living standard ,

shift to the United States people on the body .

轉嫁 到 美国 人民 身上 。

VTA0B00000000000 LIA NC3SPF0000 NE3P0M0000 LTB PP

(25) KWI(8) KWT(8) KLI(4) KMX(4) AJR00T KMX(4) KLT(4) KPT(1) KPI(8)

NE3P0M1000(16)

KLI1 LIA LTB KMX NE3P0M1O00 KLT1(21)

LI1 LT1(22)

(26) VTA0B00000000000$_4$ LI1 (24)

(30)

on American people

---

*[5] Number in parentheses refers to rule number in Table 4-5, and also indicates the precedence of rules.

*[6] The underscore indicates that one or more tag words are masked or deleted. See Table 4-5 for details.

*[7] See Section 5 for information concerning how the tags are used to choose correct English forms.

Table 4-5 — Rules Used in Sample Sentence II

Pass	Rules	Rule Number
1A	PP – KPT + PP + KPI	1
	PC – KCI + PC + KCT	2
1D	$\$\left[\begin{array}{\|c\|}\hline \mathrm{UX}\\\hline \mathrm{RX}\\\hline\end{array}\right]_1 + \mathrm{SU/IFXXX}_2 - \$\left[\begin{array}{\|c\|}\hline \mathrm{UX}\\\hline \mathrm{RX}\\\hline\end{array}\right]_1 + \mathrm{IFXXX}_2$	3
1E	LIX – KLI + LIX + KMX	4
	LTX – KMX + LTX + KLT	5
	IXXXX – KII + IXXXX + KMX	6
1G	ADXVXX$_1$ + VXXXXXXXXXXXXXXX$_2$ – KVI + VXXXXXXXXXXXXXXX$_2$ + ADXVXX$_1$ + KVT	7
	VXXXXXXXXXXXXXXX$_1$ + KLI$_2$ + LIA$_3$ – KWI + VXXXXXXXXXXXXXXX$_1$ + KWT + KLI$_2$ + LIA$_3$	8
1H	AAXXXX$_1$ + NXXXXXXXXX$_2$ – AJXXXX$_1$ + NXXXXXXXXX$_2$	9
	VTEXXXXXXXXXXXXX/NDXXXXXXXX$_1$ + $\begin{bmatrix}\mathrm{KII}\\ \mathrm{KLI}\end{bmatrix}_2$ – VTEXXXXXXXXXXXXX$_1$ + $\begin{bmatrix}\mathrm{KII}\\ \mathrm{KLI}\end{bmatrix}_2$	10
	NDXXXXXXXX$_1$ + AJXXXX/NDXXXXXXXX$_2$ – NDXXXXXXXX$_1$ + NDXXXXXXXX$_2$	11
2A	NHXXPXXXXX$_1$ + NOXXXXXXXX$_2$ – $\underline{\mathrm{NHXXPXXPXX}_1}$ + NOXXXX1XXX$_2$	12
	NEXXXXXXXX$_1$ + NXXXXXXXXX$_2$ – $\underline{\mathrm{NEXPXXXPXX}_1}$ + NXXXXX1XXX$_2$	13
	AJXXXX$_1$ + NXXXXXXXXX$_2$ – $\underline{\mathrm{AJXXXX}_1}$ + NXXXXX1XXX$_2$	14
	NDXXXXXXXX$_1$ + NDXXXXXXXX$_2$ – $\underline{\mathrm{AJXXXX}_1}$ + NDXXXX1XXX$_2$	15
	NCXXXXXXXX + NXXXXXXXXX – $\underline{\mathrm{AJXXXX}}$ + NXXXXX1XXX	16
2B	NXXXXXXXXX + PS + (KVI) + VTAXXXXXXXXXXXXX + (KVT) + NXXXXXXXXX + HM + NXXXXXXXXX – no operation	17
	KXX$_1$ + NXXXXXXXXX$_2$ + HM$_3$ + NXXXXXXXXX$_4$ + KXX$_5$ – KXX$_1$ + NXXXXX2XXX$_4$ + $\underline{\text{of + NXXXXXXOXX}_2}$ + KXX$_5$	18
	PS + VTAXXXXXXXXXXXXX/NXXXXXXXXX + NXXXXXXXXX + PS + KVI + VBAXXXXXXXXXXXXX + KVT – PS + KVI + VTAXXXXXXXXXXXXX + KVT + NXXXXXXXXX + PS + KVI + VTAXXXXXXXXXXXXX + KVT	19
	VTAXXXXXXXXXXXXX + NXXXXXXXXX + PS + KVI + VTAXXXXXXXXXXXXX + KVT + NXXXXXXXXX + PS + KVI + VTAXXXXXXXXXXXXX – KVI + VTAXXXXXXXXXXXXX + KVT + NXXXXXXXXX + PS + KVI + VTAXXXXXXXXXXXXX + KVT + NXXXXXXXXX + PS + KVI + VTAXXXXXXXXXXXXX	20

Table 4-5 — Rules Used in Sample Sentence II (Cont.)

Pass	Rules	Rule Number
3A-1	$\mathrm{KLI_1 + LIX_2 + KMX_3 + \$\overline{KXX}_4 + UP\ TO\ NEXT + KMX_5 + LTX_6 + KLT_7 + \$\overline{KLT}_8 + UP\ TO\ NEXT + \begin{bmatrix}KCI\\KPT\end{bmatrix}_9 - KLI1_1 + LIX_2 + LTX_6 + KMX_3 + \$\overline{KXX}_4 + UP\ TO\ NEXT + KLT1_7 + \$\overline{KXX}_8 + UP\ TO\ NEXT + \begin{bmatrix}KCI\\KPT\end{bmatrix}_9}$	21
3A-2	$\mathrm{LID_1 + LTA_2 - LI1_1 + LT1_2}$	22
	$\mathrm{LIA_1 + LTB_2 - LI1_1 + LT1_2}$	23
3A-3	$\mathrm{KLI1 + LXX + (LXX) + KMX + NXXXXXXXXX + KLT1 - KLI1 + LXX + \underline{(LXX) + KMX + NXXXXXXOXX} + KLT1}$	24
3D	$\mathrm{KII_1 + IXXXX_2 + KMX_3 + NXXXXXXXXX_4 + KWI_5 - KII_1 + IXXXX_2 + KMX_3 + NXXXXXXXXX_4 + KIT + KWI_5}$	25
	$\mathrm{KII_1 + IXXXX_2 + KMX_3 + NXXXXXXXXX_4 + KIT_5 - KII_1 + IXXXX_2 + \underline{KMX_3 + NXXXXXXOXX_4} + KIT_5}$	26
4	$\mathrm{VTAXXXXXXXXXXXXXXX + KVI + VTAXXXXXXXXXXXXXXX + KVT + NXXXXXXXXX + PS + KVI + VTAXXXXXXXXXXXXXXX + KVT + NXXXXXXXXX + KVI + VTAXXXXXXXXXXXXXXX + KVT + NXXXXXXXXX + HM + NXXXXXXXXX - KVI + VTAXXXXXXXXXXXXXXX + KVT + KRI + KWI + VTAXXXXXXXXXXXXXXX + KWT + NXXXXXXXXX + PS + KWI + VTAXXXXXXXXXXXXXXX + KWT + NXXXXXXXXX + PS + KWI + VTAXXXXXXXXXXXXXXX + KWT + NXXXXXXXXX + HM + NXXXXXXXXX + KRT}$	27
	$\mathrm{KRI_1 + KWI_2 + VTAXXXXXXXXXXXXXXX_3 + KWT_4 + NXXXXXXXXX_5 + PS_6 + KWI_7 + VTAXXXXXXXXXXXXXXX_8 + KWT_9 + NXXXXXXXXX_{10} + PS_{11} + KWI_{12} + VTAXXXXXXXXXXXXXXX_{13} + KWT_{14} + NXXXXXXXXX_{15} + HM_{16} + NXXXXXXXXX_{17} - NXXXXXRXXX_{17} + \underline{of + VTAXXXXXXXXXXXPXXX_3 + NXXXXXXOXX_5 + PS_6 + VTAXXXXXXXXXXXPXXX_8 + NXXXXXXOXX_{10} + PS_{11} + VTAXXXXXXXXXXXPXXX_{12} + NXXXXXXOXX_{15}}}$	28
	$\mathrm{KII_1 + IXRRX_2 + KIT_3 + KWI_4 + VXXXXXXXXXXXXXXXXX_5 + KWT_6 + KLI_7 - KWI_4 + VXXXXXXXXXXXXXXXXX_5 + KWT_6 + KII_1 + IXRRX_2 + KIT_3 + KLI_7}$	29
	$\mathrm{KLIX + LXX + KLTX + \begin{bmatrix}KCI\\KPT\end{bmatrix} - \underline{KLIX + LXX + KLTX} + \begin{bmatrix}KCI\\KPT\end{bmatrix}}$	30
	$\mathrm{KII + IFXXX + KIT - NXXXXXXXXX}$	31
5	$\mathrm{NX3SXXXXXX + VTAXXXXXXXXXXXXXXX + NXXXXXXXXX + ,\ + VTAXXXXXXXXXXXXXXX + NXXXXXXXXX - NX3SXXXSXX + VTAXXXXXXXXXXX3SXX + NXXXXXXOXX + ,\ + VTAXXXXXXXXXXX3SXX + NXXXXXXOXX}$	32*

*This is a combination of several rules. For details, see rules in pass 5 of rule file.

5. ENGLISH WORD INFLECTIONS

In English output the final forms of nouns, adjectives, adverbs, and verbs are composed by adding auxiliaries and endings to the English words in the translation section of the dictionary entry cards. The form that these words are to assume in translation is determined by the English tags. The auxiliaries and endings to be added to each English word are listed in the tables described below.

English words in the translation section are listed in stem form. The end of the stem is marked by a 0. Directly following the 0 are tags that designate the tables in which auxiliaries and endings are to be looked up for addition to the English stem. These tags are distinguished by the initial letters N (noun), A (adjective and/or adverb), and V (verb). Following the initial letters are two or three letters and/or numbers designating the appropriate tables. In the event that a word is too irregular to be declined according to the tables, its separate forms will be listed in a column in the translation section of the dictionary entry card.

This section contains descriptions of tags and tables to be used for English stems in the nominal, adjectival-adverbial, and verbal categories. Appendix C contains tables showing the forms verbals will take as the relevant variables in the English grammatic tags are changed.

5.1 NOUN INFLECTIONS

Nouns have four possible forms: singular nominative, singular possessive, plural nominative, and plural possessive. Singular and plural forms are determined by S or P in the fourth tag of the nominal tag word. Nominative and possessive forms are determined by X or P in the eighth tag. These four forms of the noun are made by adding endings to the stem form of the translation.

Three table designators—a letter, a number, and a letter—are written directly after the 0N that follows the stem. The table designators indicate which endings are to be added to the stem and where the additions are to be made. The first table designator specifies what endings are to be added to form the singular and plural. The second table designator indicates which word in the stem is to be pluralized, and the third table designator tells how the possessive is created. For example, the stem and table designators DOG0NS0A would take the following forms with the four possible variations of the nominal tag word:

1. NXSXXXXXXX (singular nominative): dog
2. NXSXXXXPXX (singular possessive): dog's
3. NXPXXXXXXX (plural nominative): dogs
4. NXPXXXXPXX (plural possessive): dogs'

5.1.1 Singular and Plural Table Designators

The first table designator is a letter indicating the manners of forming singular and plural nouns shown in Table 5-1.

5.1.2 Plural Position Table Designators

The numbers 0, 1, 2, and 3 indicate which word in the stem is to be pluralized, as shown in Table 5-2.

5.1.3 Possessive Table Designators

The letters A, B, and C designate the three tables of possessive endings, shown in Table 5-3, that are always added to the end of the singular or plural word.

5.1.4 Irregular Nouns

Irregular nouns that form the plural in a manner other than that covered in Table 5-1 will be manually entered in the translation section of the dictionary entry card in the following order:

1. NXSXXXXXXX (singular nominative): person
2. NXSXXXXPXX (singular possessive): person's
3. NXPXXXXXXX (plural nominative): people
4. NXPXXXXPXX (plural possessive): people's

5.1.5 Examples

The following examples show the English inflection translation process:

1. REVOLUTIONARY BASE0NS0A: NXSXXXXXXX (revolutionary base)
   NXSXXXXPXX (revolutionary base's)
   NXPXXXXXXX (revolutionary bases)
   NXPXXXXPXX (revolutionary bases')

Table 5-1 — Singular and Plural Table Designators

Table Designator	Table of Endings	Examples
0	No table	goose, geese (listed separately)
I	S—no plural	real estate
P	P—no singular	Phillippine Islands
S	S—stem + 0 (no change)	dog
	P—stem + -s	dogs
E	S—stem + 0	trench
	P—stem + -es	trenches
Y	S—stem + -y	fly (note stem form "fl")
	P—stem + -ies	flies
A	S—stem + -an	workman (note stem form "workm")
	P—stem + -en	workmen
F	S—stem + -f	shelf (note stem form "shel")
	P—stem + -ves	shelves
N	S—stem + 0	sheep
	P—stem + 0	sheep

Table 5-2 — Plural Position Table Designators

Number	Word Pluralized	Examples
0	last	Communist Parties
1	first	friends of peace
2	second	international agreements on human rights
3	first and third	fathers and sons

Table 5-3 — Possessive Table Designators

Letter	Word Made Possessive	Examples
A	S—Singular nominative + -'s	dog's
	P—Plural nominative + '	dogs'
B	S—Singular nominative + -'s	sheep's
	P—Plural nominative + -'s	sheep's
C	S—Singular nominative + '	puppet of the United States'
	P—Plural nominative + '	puppets of the United States'

2. M OF LETTERSONA1C: NXSXXXXXXX (man of letters)
NXSXXXXPXX (man of letters')
NXPXXXXXXX (men of letters)
NXPXXXXPXX (men of letters')

3. PROLETARIATONI0B: NXSXXXXXXX (proletariat)
NXSXXXXPXX (proletariat's)

4. TROOPSONP0C: NXPXXXXXXX (troops)
NXPXXXXPXX (troops')

5.2 ADJECTIVE AND ADVERB AUXILIARIES AND INFLECTIONS

The adjectival and adverbial forms (AJ and AD) normally have regular, comparative, and superlative forms (R, C, and S) made by adding auxiliaries and/or endings to the stem form of the English translation. The additions to a given stem will be specified by the addition of two table designators written directly following the 0A at the end of the stem. The first table designator specifies how the different forms of the adjective are created. The second table designator specifies how the different forms of the adverb are made from the adjective stem.

5.2.1 Adjective Formation

Table 5-4 shows the letters that designate common pattern of adjective formation.

5.2.2 Adverb Formation

Adverbs normally conform to the ending tables shown in Table 5-5.

5.2.3 Examples

1. PEACEFUL0AML: AJR (peaceful)
AJC (more peaceful)
AJS (most peaceful)
ADR (peacefully)
ADC (more peacefully)
ADS (most peacefully)

2. CLUMS0AYI: AJR (clumsy)
AJC (clumsier)
AJS (clumsiest)

Table 5-4 — Adjective Table Designators

Table Designator	Adjectival Ending Table	Examples
0	No table	No adjectival form
M	R—stem + 0	progressive
	C—more + stem	more progressive
	S—most + stem	most progressive
L	R—stem + -le	capable (note stem "capab")
	C—more + stem + -le	more capable
	S—most + stem + -le	most capable
R	R—stem + 0	white
	C—stem + -r	whiter
	S—stem + -st	whitest
E	R—stem + 0	fast
	C—stem + -er	faster
	S—stem + -est	fastest
Y	R—stem + -y	funny
	C—stem + -ier	funnier
	S—stem + -iest	funniest

Table 5-5 — Adverb Table Designators

Table Designator	Adverbial Ending Table	Examples
0	No table	No adverbial form
R	R—stem + 0	late
	C—stem + -r	later
	S—stem + -st	latest
E	R—stem + 0	soon
	C—stem + -er	sooner
	S—stem + -est	soonest
Y	R—stem + -y	early (note stem form "earl")
	C—stem + -ier	earlier
	S—stem + -iest	earliest
A	R—stem + -ally	scientifically
	C—more + stem + -ally	more scientifically
	S—most + stem + -ally	most scientifically
I	R—stem + -ily	hungrily
	C—more + stem + -ily	more hungrily
	S—most + stem + -ily	most hungrily
L	R—stem + -ly	treacherously
	C—more + stem + -ly	more treacherously
	S—most + stem + -ly	most treacherously

ADR (clumsily)
ADC (more clumsily)
ADS (most clumsily)

3. COUNTERREVOLUTIONARY0AM0: AJR (counterrevolutionary)
AJC (more counterrevolutionary)
AJS (most counterrevolutionary)

5.3 VERBS

All English verb forms may be made by adding various kinds of auxiliaries and endings to a basic verb stem. These additions may be listed in table form with appropriate code letters so that they can be looked up and combined in accordance with the person, number, tense aspect, and mode as specified by positions 12 through 16 in the English tag. The tables of verb components are listed below under five headings: (1) verb stems and endings, (2) forms of the verb "to be," (3) auxiliary verbs, (4) infinitive auxiliaries, and (5) postmodal auxiliaries.

In the following verb tables these letters are used to designate the various kinds of verb components:

1. V—verb stem plus ending
2. B—forms of the verb "to be"
3. A—auxiliary verbs (positive)
4. D—auxiliary verbs (negative)
5. I—infinitive auxiliaries (positive and negative)
6. G—postmodal auxiliaries (positive and negative)

5.3.1 Verb Stems Plus Endings

Each verb stem has the six basic forms shown in Table 5-6. The manner in which a verb stem assumes these forms is shown by three table designators written after the 0V at the end of the stem. The first table designator indicates formation of forms V1 through V4. The second determines forms V5 and V6. The third table designator specifies which word in the verb stem is to change form.

Forms V1 Through V4. The letters S, E, C, V, and Y indicate the most common forms of stem change for English verbs, as shown in Table 5-7.

Forms V5 and V6. Weak or regular verbs commonly form the past and past participle in the manner indicated by the table designators E, C, Y, and O in Table 5-8.

Table 5-6 — The Six Basic Verb Stem Forms

Number	Form	Examples		
0	Stem	kill	go	lov
V1	Present, 1S	kill	go	love
V2	Present, 3S	kills	goes	loves
V3	Present participle	killing	going	loving
V4	Infinitive	to kill	to go	to love
V5	Past	killed	went	loved
V6	Past participle	killed	gone	loved

Table 5-7 — The Most Common Forms of Stem Change for English Verbs

Table Designator	Verb Ending	
S	V1—stem + 0	spell
	V2—stem + -s	spells
	V3—stem + -ing	spelling
	V4—to + stem + 0	to spell
E	V1—stem + 0	push
	V2—stem + -es	pushes
	V3—stem + -ing	pushing
	V4—to + stem + 0	to push
C	V1—stem + 0	grin
	V2—stem + -s	grins
	V3—stem + consonant + -ing	grinning (note doubled consonant)
	V4—to + stem + 0	to grin
V	V1—stem + -e	love (note that stem lacks vowel "e")
	V2—stem + -es	loves
	V3—stem + -ing	loving
	V4—to + stem + -e	to love
Y	V1—stem + -y	try
	V2—stem + -ies	tries
	V3—stem + -ying	trying
	V4—to + stem + -y	to try

Table 5-8 — Forms of Stem Changes for Weak or Regular Verbs

Table Designator	Verb Ending	Examples
E	V5—stem + -ed	spelled
	V6—stem + ed	spelled
C	V5—stem + consonant + -ed	grinned
	V6—stem + consonant + -ed	grinned
Y	V5—stem + -ied	tried
	V6—stem + -ied	tried
O	V5—stem + 0	put
	V6—stem + 0	put

Verb Stem Position Indicator. A number is used to specify which word in a verb stem is to change form. A 0 indicates that the ending is added to the last word in the stem, a 1 that it is added to the first word (if the stem has more than one word), a 2 that is added to the second word, etc.

Irregular Verbs. In the case of verbs that are too irregular for the past and past participle forms to be predicted by the above tables, these forms will simply be listed, in the translation section of the dictionary entry card, preceded by V5 and V6. Thus a word like "see" would show SEE0VS00 in the translation section, directly beneath which would be listed V5SAW and V6SEEN.

Examples. The examples in Table 5-9 show the stem and table designators above the six verb forms specified by them.

5.3.2 Forms of the Verb "To Be"

The verb "to be" is the most irregular verb in English. Consequently, its various forms cannot be predicted by a stem and table of endings. A total of nine different spellings must be listed separately (Table 5-10).

Forms of the verb "to be" not shown in Table 5-10 may be made from auxiliary verbs, infinitive auxiliaries, and postmodal auxiliaries with or without the addition of the forms shown.

When the verb "to be" is part of a verb stem listed in the translation section of a dictionary entry card, it will appear as the letter B. For example, the stem of the verb "to be stupid" will appear as B STUPID0VB01. Note that for the verb "to be" the first table designator will show the letter B and the second will always be 0 (because the fifth and sixth verb forms are already included in the B table).

5.3.3 Auxiliary Verbs

All auxiliary verbs help to form all tenses except the present and past and are the same for all verbs except the verb "to be." In the case of "to be," the same set of auxiliaries may be utilized, but in slightly different tenses. Tables 5-11 through 5-14 show all possible auxiliary forms in both positive and negative form. Those forms that do not have a number do not occur in the present Chinese to English MT system and are included here for information purposes only.

Positive Auxiliary Verbs. There are 18 numbered auxiliary verbs that can occur in English output, as seen in Table 5-11.

Negative Auxiliary Verbs. The number of negative auxiliaries is three greater than the number of positive auxiliaries because of the necessity to negate the present and past forms of the

Table 5-9 — Examples of Verb Stems and Ending Formations

0	WORK0VSE0	FL0VY00	ORGANIZ0VVE0	OCCUP BY FORCE0VYY1
V1	work	fly	organize	occupy by force
V2	works	flies	organizes	occupies by force
V3	working	flying	organizing	occupying by force
V4	to work	to fly	to organize	to occupy by force
V5	worked	0 (flew)	organized	occupied by force
V6	worked	0 (flown)	organized	occupied by force

Table 5-10 — Nine Forms of the Verb "To Be"

Number	Name	Examples
B1	First person singular (present)	am
B2	Second person singular (present)	are
B3	Third person singular (present)	is
B4	First person singular (past)	was
B5	Second person singular (past)	were
B6	Infinitive	to be
B7	Infinitive minus "to"	be
B8	Present participle	being
B9	Past participle	been

Table 5-11 — Auxiliary Verbs That Can Occur in English Output

Number	Name	Examples
A1	First person singular present perfect active	have
A2	Third person singular present perfect active	has
A3	First person singular past perfect active	had
A4	First person singular future active	will
A5	First person singular future perfect active	will have
A6	First person singular present progressive active	am
A7	Second person singular present progressive active	are
A8	Third person singular present progressive active	is
A9	First person singular present perfect progressive active	have been
A10	Third person singular present perfect progressive active	has been
A11	First person singular past progressive active	was
A12	Second person singular past progressive active	were
A13	First person singular past perfect progressive active	had been
A14	First person singular future progressive active	will be
A15	First person singular future perfect progressive active	will have been
A16	First person singular present progressive passive	am being
A17	Second person singular present progressive passive	are being
A18	Third person singular present progressive passive	is being
A__	First person singular present perfect progressive passive	(have been being)
A__	Third person singular present perfect progressive passive	(has been being)
A__	First person singular past progressive passive	(was being)
A__	Second person singular past progressive passive	(were being)
A__	First person singular past perfect progressive passive	(had been being)
A__	First person singular future progressive passive	(will be being)
A__	First person singular future perfect progressive passive	(will have been being)

verb (V1, V2, and V5) that require no auxiliaries in the present. Note (Table 5-12) that some of the negative auxiliaries (e.g., D10, D11, and D12) constitute the negative forms of the verb "to be."

5.3.4 Infinitive Auxiliaries

Infinitive forms are made by the addition of an appropriate auxiliary to a V or B form. Both positive and negative forms are listed in Table 5-13. Ten of these forms may appear in English output.

5.3.5 Postmodal Auxiliaries

Since verbs following modals (could, must, etc.) take auxiliaries that are really infinitives minus the initial "to," a special table of auxiliaries is required. Nine positive and negative forms are listed in Table 5-14.

5.4 NOUN, ADJECTIVE-ADVERB, AND VERB TABLE DESIGNATORS

Table 5-15 is provided as a brief guide to be used by linguists filling out dictionary entry cards.

Table 5-12 — Negative Auxiliary Verbs

Number	Name	Examples
D1	First person singular present active	do not
D2	Third person singular present active	does not
D3	First person singular present perfect active	have not
D4	Third person singular present perfect active	has not
D5	First person singular past active	did not
D6	First person singular past perfect active	had not
D7	First person singular future active	will not
D8	First person singular future perfect active	will not have
D9	First person singular present progressive active	am not
D10	Second person singular present progressive active	are not
D11	Third person singular present progressive active	is not
D12	First person singular present perfect progressive active	have not been
D13	Third person singular present perfect progressive active	has not been
D14	First person singular past progressive active	was not
D15	Third person singular past progressive active	were not
D16	First person singular past perfect progressive active	had not been
D17	First person singular future progressive active	will not be
D18	First person singular future perfect active	will not have been
D19	First person singular present progressive passive	am not being
D20	Second person singular present progressive passive	are not being
D21	Third person singular present progressive passive	is not being
D__	First person singular present perfect progressive passive	(have not been being)
D__	Third person singular present perfect progressive passive	(has not been being)
D__	First person singular past progressive passive	(was not being)
D__	Second person singular past progressive passive	(were not being)
D__	First person singular past perfect progressive passive	(had not been being)
D__	First person singular future progressive passive	(will not be being)
D__	First person singular future perfect progressive passive	(will not have been being)

Table 5-13 — Positive and Negative Infinitive Auxiliaries

Number	Name	Formation	Examples
I1	Present active	V4	to see
I2	Present perfect active	to have + V6	to have seen
I3	Present progressive active	to be + V3	to be seeing
—	Present perfect progressive active	to have been + V3	to have been seeing
I4	Present passive	to be + V6	to be seen
I5	Present perfect passive	to have been + V6	to have been seen
—	Present progressive passive	to be being + V6	to be being seen
—	Present perfect progressive passive	to have been being + V6	to have been being seen
I6	Present active	to not + V1	to not see
I7	Present perfect active	to not have + V6	to not have seen
I8	Present progressive active	to not be + V3	to not be seeing
—	Present perfect progressive	(to not have been) + V3	to not have been seeing
I9	Present passive	to not be + V6	to not be seen
I10	Present perfect passive	to not have been + V6	to not have been seen
—	Present progressive passive	(to not have been being) + V6	to not have been being seen

Table 5-14 — Positive and Negative Postmodal Auxiliaries

Number	Name	Examples
G1	Present passive	be
G2	Present progressive passive	be being
G3	Present perfect active	have
G4	Present perfect passive	have been
G5	Present active (negative)	not
G6	Present passive (negative)	not be
G7	Present progressive passive (negative)	not be being
G8	Present perfect active (negative)	not have
G9	Present perfect passive (negative)	not have been

Table 5-15 — Table of Noun, Verb, and Adjective-Adverb Table Designators

Nouns—N

Plural Endings	Word Indicator	Possessive Endings
0 (no table)	0 (last word)	A (S: + -'s)
I (no plural)	1 (first word)	(P: + ')
P (no singular)	2 (second word)	B (S: + -'s)
S (add -s)	3 (first and third words)	(P: + -'s)
E (add -es)		C (S: + ')
Y (-y becomes -ies)		(P: + ')
A (-an becomes -en)		
F (-f becomes -ves)		
N (no change)		

Verbs—V

Forms 1-4	Forms 5 and 6	Word Indicator
0 (no table)	0 (no table)	0 (last word)
B (all forms of "to be")	E (stem + -ed)	1 (first word)
S (3S adds -s)	C (+ consonant + -ed)	2 (second word)
E (3S adds -es)	Y (stem + -ied)	3 (third word)
C (consonant + -ing)	O (stem + 0)	
V (1S adds -e to stem)		
Y (1S adds -y to stem)		

Adjective-Adverbs—A

Adjective Endings	Adverb Endings
0 (no table)	0 (no table)
M (+ more and most)	R (+ -r and st)
L (+ -le and more and most)	E (+ -er and -est)
R (+ -r and -st)	Y (+ -y and -ier and -iest)
E (+ -er and -est)	A (+ -ally and more and most)
Y (+ -y and -ier and -iest)	I (+ -ily and more and most)
	L (+ -ly and more and most)

6. CONCLUSIONS AND RECOMMENDATIONS

The first year of linguistic studies for Chinese to English machine translation has produced some very significant results in establishing a general linguistic processing system for Chinese to English MT. The tasks accomplished through the efforts of the linguistic and programming groups allow this study to be effective for a programmed system for further refinement and development.

The tasks accomplished during this period are as follows:

1. A general linguistic MT system, beginning from the Chinese input by Chicoder to the output of English words, detailing specific linguistic passes and the operations necessary to reach the goal of the target language output, was developed.

2. Morphological analyses resulting in the setting up of table formats for word classes and their subtags were carried out. Within these word classes and their subtags, flexibility is allowed for expansion and/or revision. Lexicographic entries can be made on a systematic production basis. As a result, 15,000 Chinese input entries and 30,000 English output entries plus input and output entries of general vocabulary based on linguistic analysis are put on magnetic tape.

3. Syntactical analyses were performed, with the following results:
    - The establishment of linguistic rules for resolving ambiguities, phrase and structure recognition, word reordering, etc.
    - The organization of these linguistic rules into a system wherein certain operations and items are sequenced for logic operations
    - The writing of over 1,000 linguistic rules in symbolic notation to process linguistic structures discovered in the linguistic analysis.

4. General English output with all the inflection forms for nominals, verbals, and adjectival adverbials set up in table format for lexicographic and programming operations.

5. Sentence structures and patterns were analyzed for general format for subject to verb to object linking.
6. Testing of linguistic rules by machine was initiated and is a continuing process to aid the progress of linguistic research.

For a fine system of Chinese to English machine translation, refinement and development of the basic system that we now have are necessary. It is therefore proposed that future work concentrate on the refinement and amplification of this system. General items of work to be considered for the future should include the following:

1. Study of scientific texts in the source language to allow research for any variances in grammatical structures and for any elaboration of these structures

2. Expansion of morphological tag classification and linguistic rules based on study of texts

3. Further analysis of sentence patterns to expand detailed operations of linking of major phrases

4. Addition of lexicographic entries for linguistic study purposes

5. Refinement and further development of the linguistic processing system

6. Continuous computer testing of linguistic rules, linguistic passes, and correlation of one pass to another, as well as special testing of English output with all the inflection forms, for the purpose of testing the validity of the linguistic systems.

Specifically, the details for the proposed work can be divided into two categories, morphology and syntax. Although these details are considered to be two categories, in the actual process of linguistic analysis both will be utilized simultaneously to produce the best results.

6.1 MORPHOLOGY

Future morphological work should include the following:

1. Addition of lexicographic entries of scientific terminology should be made either from one particular field or from general scientific texts. These entries can be utilized for testing purposes.
2. Addition of lexicographic entries should include all common verbs, nouns, and idiomatic expressions to ensure proper application of the linguistic rules in the processing system.
3. Proverbs that are quoted frequently should be entered in lexicographic entries and classified accordingly.

4. Further analysis of the verbs should include:

   - Information concerning the possible kind of subject each verb may take. At least a major division between human verbs and those verbs that are related only to abstract fields such as science, politics, and economics should be made.
   - Information concerning the relationship between certain verbs and the object. If the relationship is so close that it influences the English equivalent words, tag modifications and addition of linguistic rules will be needed.

5. Further adjectival-adverbial analysis concerning information of certain adjectives that cause the verb "to be" to be omitted is necessary. Addition of linguistic rules should be studied for adverbs and supplemented accordingly.
6. Since the present collocative tag word does not carry the information concerning its adjectival form, this information should be added.
7. Single character verbs should be further analyzed.
8. Special tags should be added in the verb tag word for verbs whose subjects are omitted.
9. The auxiliary English verb table should be set up and an appropriate tag system made for the G class so that "can" and "may" can be conjugated.
10. New classes should be created for words that have not been classified in the present system.

6.2 SYNTAX

Future syntactical work should include the following:

1. A study should be made on the finding of proper names so that any proper name will be found and tagged accordingly before the sentence enters the grammatic analysis processing phase in the scheme.
2. Since the English sentence structure seldom omits the subject except in imperative sentences, and since the Chinese sentence often omits the subject after it is mentioned once, a study should be made so that the resulting translation will be natural English.
3. More study should be made on collocative structures, connominal phrases, and noun phrase-relative clauses so that these structures can be safely extended over commas.
4. Studies should be made on collocative structures, connominal phrases, and noun

phrase-relative clauses so that new patterns may be discovered and more natural translations may be derived for these structures.

5. **Syntactical study should be made for comparative structures and interrogative sentences. Appropriate tag words, linguistic rules, and English equivalent structures should be made for this study.**
6. **Studies should be made for machine recognition of main clauses and subordinate clauses.**
7. **Addition of English articles for smoother translation should be studied.**
8. **Concordances of problem words and single characters should be made.**
9. **More parsing of texts is recommended.**

Past experience has demonstrated that greater progress can be made in this area of linguistic research if the linguistic analysis is supplemented by computer testing. It is therefore suggested that program testing of rules, passes, and English output be continued to test the validity and the systematic operation of the linguistic rules.

7. PROJECT PERSONNEL

The linguistic studies work on this contract was performed by the linguistic researchers with the cooperation of the Computer Sciences Department. The work was done in the Digital Data Systems Division of Itek Corporation.

The Itek linguistic research personnel are:

T. C. Lee
J. L. Wong
H. T. Wang
E. L Farmer
J. M. Polachek
S. C. Hwang
P. H. Ho
J. R. Lim

The Itek computer sciences personnel are:

C R. Burgess
S. C. Yang
C. E. Hurlburt
B. Howley
T. R. Cullen

Dr. Lien-sheng Yang, Harvard University Professor of Far Eastern languages, acted as consultant throughout this study.

8. BIBLIOGRAPHY

1. Bach, E., "An Introduction to Transformational Grammars," Holt, Rinehart and Winston, Inc., New York, 1964.
2. Bloomfield, L., "Language," Holt, Rinehart and Winston, Inc., New York, 1962.
3. Chao, Y. R. and Yang, L. S., "Concise Dictionary of Spoken Chinese," Harvard University Press, Cambridge, Massachusetts, 1952.
4. Chao, Y. R., "Mandarin Primer," Harvard University Press, Cambridge, Massachusetts, 1952.
5. "Chinese-English Dictionary of Modern Communist Chinese Usage," Joint Publications Research Service, U.S. Department of Commerce, Washington, D.C., 1963.
6. Chomsky, N., "Syntactic Structures," Mouton and Company, The Hague, The Netherlands, 1957.
7. Dougherty, C-Y, "The Lexeme 'De' as a Syntactic Marker," University of California Press, Berkeley, California, 1964.
8. Dougherty, C-Y and Martin, S. E., "Chinese Syntactic Rules for Machine Translation," University of California Press, Berkeley, California, 1964.
9. Fillmore, C. J., Meyers, L. F., and Wang, W. S-Y, "Project on Linguistics Analysis, Report No. 3," Ohio State University Press, Columbus, Ohio, 1963.
10. Final Report on Chinese-English Machine Translation Research, International Business Machines Corp., AF 30(602)-2479, ADC-TDR-63-303 (1963).
11. Gleason, H. A., "An Introduction to Descriptive Linguistics," Henry Holt and Company, New York, 1958.
12. Harris, Z. S., "Structural Linguistics," University of Chicago Press, Chicago, Illinois, 1951.
13. Hockett, C. F., "A Course in Modern Linguistics," The MacMillan Company, New York, 1959.
14. HongQi Magazine, HongQi Magazine Editing Committee, Peking, China (1963 through 1965).
15. Joos, M. (Ed.), "Readings in Linguistics," 3rd ed., American Council of Learned Societies, New York, 1963.

16. King, G. W. and Chang, H. W., Machine Translation of Chinese, Scientific American, 208(6):124-135 (June 1963).

17. Kittredge, G. L. and Farley, F. E., "An Advanced English Grammar," Ginn and Company, Boston, Massachusetts, 1913.

18. Locke, W. N. and Booth, A. D. (Eds.), "Machine Translation of Languages," Technology Press of the Massachusetts Institute of Technology and John Wiley and Sons, Inc., New York, 1955.

19. Marckward, A. H. (Ed.), "Studies in Language and Linguistics," University of Michigan Press, Ann Arbor, Michigan, 1964.

20. Mathews' Chinese-English Dictionary," Harvard University Press, Cambridge, Massachusetts, 1963.

21. "Modern Chinese Reader," parts I and II, Epoch Publishing House, Peking, China, 1962.

22. Reference Manual—Chicoder Code, Itek 1023-1 (20 July 1964).

23. Ren Min Ri Bao (People's Daily), Peking, China (1964 and 1965).

24. "Roget's International Thesaurus," Thomas Y. Crowell Company, New York, 1964.

25. Wang, W. S-Y, "Project on Linguistics Analysis, Report No. 4," Ohio State University Press, Columbus, Ohio, 1963.

26. "Webster's New International Dictionary," 2nd ed., G&C Merriam Company, Springfield, Massachusetts, 1959.

Appendix A

EXPLANATION OF SYMBOLS FOR LINGUISTIC RULES

BASIC RULE FORMAT

$$A_1 + B/P_2 + K \sim \substack{3\\3} \rightarrow C_1 + \underline{D + B_2}$$

The above is the basic linguistic rule format, and all other formats are derived from it. The following are the explanations of the symbols used.

1. +—the plus sign represents the boundary between grammatical tag words belonging to different Chinese words, or morphemes as segmented and tagged according to the dictionary entries.

2. Letters—the first letter after every plus sign or at the beginning of each rule represents the first tag letter of each tag word. For example, +N represents the words "followed by a nominal"; +A represents "followed by an adjectival adverbial." If no more letters follow this first letter or if this letter is followed by another +, then it is understood that this first letter represents the entire tag word of the class it indicates, with the rest of the tag word filled by X, to be explained later in this Appendix. Therefore:

+ A + is the same as + AXXXXX +

+ N + is the same as + NXXXXXXXXX +

If the first letter is followed by another letter with no number above it, then the two letters represent the first two tag letters of the tag word. Therefore:

+ AJ + is the same as + AJXXXX +

+ EN + is the same as + ENXX +

If there is a number above the letter following the first letter, the number indicates the tag number of the letter below it. Therefore:

$N\overset{10}{V}$ represents NXXXXXXXXV

The result of combining the above two points is:

$NT\overset{9}{D}$ represents NTXXXXXXDX

$VTA\overset{8}{1}$ represents VTAXXXX1XXXXXXXX

(For the exact length of every tag word, see tables in Section 3.)

3. ~—the tilde is used when the rest of the tag word is occupied by X's. The superscript to a tilde indicates the number of the last tag to be filled by X. Therefore:

$VTA \sim^{16}$ means VTAXXXXXXXXXXXXX

$V \sim^{11}$ V3SPA means VXXXXXXXXXXV3SPA

4. /—if a letter or a group of letters is followed by a slash and another letter or a group of letters, it is understood that the actual Chinese word owning these two groups of letters falls into two grammatical classes represented by these two tag words, i.e., that this word is a grammatic ambiguity. Some Chinese words fall into three or four grammatic classes. The different tag words indicating the classes are divided by slashes. The order of the classes thus represented is: R/A/L/S/I/J/V/E/N/U/G/O/H/P. Therefore, if a word falls into classes S and A, the symbol will be A/S or AXXXXX/SX. If a word falls into classes V, L, J, and N, the symbol will be L/J/V/N or LXX/JX/VXXXXXXXXXXXXXXX/NXXXXXXXXXX.

5. Number—the subscripts to the letters of the alphabet represent the sequential number of the word stream, $A_1 + B_2 + K_3$ means that tag word A is immediately followed by tag word B, which is immediately followed by tag word K.

6. ______ — when a letter or a series of letters is underlined, it means that the tag words represented by the letter are masked. Masking means that the underscored tag words will be treated as if they were not there in subsequent linguistic passes, but they will be translated accordingly in the output. This is the technique by which the head of a noun phrase is shown to be the representative of the noun phrase, etc. Masking should by no means be confused with deleting.

7. → — the material on the left side of the arrow is the argument, and that on the right is the function. Whatever is on the left is replaced by whatever is on the right.

$$A_1 + B/P_2 + K_3 \rightarrow C_1 + \underline{D + B_2}$$

The above expression is read as follows: When tag word A is immediately followed by the ambiguity tag word B/P, and B/P is immediately followed by tag word K, tag word A is changed to tag word C, ambiguity tag word B/P is solved into tag word B with tag word P deleted, tag word D is inserted between the original tag words A and B/P, tag word K is deleted, and tag words D and B are masked.

OTHER SYMBOLS

1. X— X represents any letter or number in its position. Therefore, NX will represent all the noun types, and AXC will represent all the adjectivals and adverbials that are in the comparative degree.

2. ( )— when parentheses appear in a rule, it means that whatever is included in the parentheses may be found in that position. Therefore, A + (B + D) + C represents the two sequences:

A + B + D + C

A + C

A + (B) + C + (D) represents the four sequences:

A + B + C

A + C

A + B + C + D

A + C + D

3. [ ]— when brackets appear in a rule, it means that there is a choice between or among whatever is included. Therefore, $A + \begin{bmatrix} B \\ C \end{bmatrix} + D$ represents the two sequences:

A + B + D

A + C + D

The symbol "$" is explained in Table A-1.

Table A-1 — The Symbol "$"

$Z + \$ + \text{UP TO NEXT} + Y$	Start looking for Y from Z, stop looking for Y when meeting KCI or KPT (rule does not apply): if Y is found, rule applies; if not, rule does not apply.
$Z + \overset{P}{\$} + \text{UP TO NEXT} + Y$	Start looking for Y from Z, stop looking for Y when meeting KPT (rule does not apply): if Y is found, rule applies; if not, rule does not apply.
$Z + \$\overline{W} + \text{UP TO NEXT} + Y$	Start looking for Y from Z, stop looking for Y when meeting KCI, KPT or W (rule does not apply): if Y is found, rule applies; if not, rule does not apply.
$Z + \overset{P}{\$}\overline{W} + \text{UP TO NEXT} + Y$	Start looking for Y from Z, stop looking for Y when meeting KPT or W (rule does not apply): if Y is found, rule applies; if not, rule does not apply.
$\$\overline{W} + \text{UP TO NEXT} + Y$	Start looking for Y from preceding KCT or KPI, stop looking for Y when meeting KCI, KPT, or W (rule does not apply): if Y is found, rule applies; if not, rule does not apply.
$\overset{P}{\$}\overline{W} + \text{UP TO NEXT} + Y$	Start looking for Y from KPI, stop looking for Y when meeting KPT or W (rule does not apply): if Y is found, rule applies; if not, rule does not apply.
$\$\overline{W} + Y$	If W immediately precedes Y, rule does not apply; otherwise, rule applies.
$Y + \$\overline{W}$ (immediate)	If W immediately follows Y, rule does not apply; otherwise, rule applies.

KXX

Any tag word that starts with a K does not belong to the classification tag word of the dictionary entries. K tag words are used to indicate the beginning or ending of specific phrases or clauses. They are inserted through linguistic rules rather than through dictionary lookup. When a certain environment is found, certain phrase or structure indicators are inserted. Table A-2 gives a list of phrase and structure indicators used.

Table A-2 — Phrase and Structure Indicators

Indicator	Meaning
KCI	Beginning of a comma
KCT	Ending of a comma
KPI	Beginning of a sentence
KPT	Ending of a sentence
KVI	Beginning of major verbal phrase
KVT	Ending of major verbal phrase
KWI	Beginning of minor verbal phrase
KWT	Ending of minor verbal phrase
KII	Beginning of connominal phrase
KIT	Ending of connominal phrase
KLI	Beginning of collocative structure
KLT	Ending of collocative structure
KRI	Beginning of noun phrase-relative clause
KRT	Ending of noun phrase-relative clause
KMX	Minor phrase

NOTE: Except for KMX, the KXX's usually go in pairs.

Appendix B

LISTINGS OF COMPUTER EXPERIMENTATION

This Appendix presents the listing of computer processing from Chinese to English, including Chinese text (both in Chinese and Chicodes), dictionary I, output of pass 1, dictionary II, output of pass 2, dictionary III, output of pass 3, and final output of English after the inflection process.

CHINESE TEXT IN CHINESE AND CHICODES

1v52/fh14/vf11/ss11/e911/1v11/ss11/8y11/fr11/ss14/e911/1v11/4j12/bs14/dp11/uh12/yh11/

新的一年，一九六五年，已经来临了

1v12/ci11/qs11/oo11/4i11/5e11/km13/qw12/xu11/vf11/wg13/5y21/xg11/ch11/ak12/sk11/8o11/

我国人民怀着胜利的喜悦和争取更大

qw12/xu11/vf11/rg14/1511/1v11/jh21/ht15/jk11/fh14/vf11/ss11/e911/vf11/su11/dp11/1v12/

胜利的信心，迎接这新的一年的到来.

DICTIONARY I

Chicodes	Pseudo-Chicodes and Chinese Grammatic Tags
ch11/ak12/	vaaaaa/vta0002000000000)
ci11/	naaa00/nm1spm0000)
dp11/uh12/	vaabaa/vi00000000000000)
e911/	naab00/nt3s0n00u0)
fh14/	aaaa00/aarv0t)
fr11/	uaaa00/up)
jh21/ht15/	vaacaa/vta0000000000000)
jk11/	raaa00/rs)
km13/	eaaa00/ept0)
oo11/4i11/	naacaa/ne3p0m0000)

qs11/	naad00/nc3s0n0000)
qw12/xu11/	caaaaa/vi00000000000000)nd3s0n0000)
rg14/1511/	naaeaa/nd3s0ny000)
sk11/	aaab00/adcr0t)
ss11/	uaab00/us)
ss14/	uaac00/up)
su11/dp11/	vaadaa/vi00000000000000)
vf11/	haaa00/hm)
wg13/5y21/	aaacaa/ajr00t)
xg11/	baaa00/iaaan)jc)
yh11/	eaab00/ette)
1v11/	paaa00/pc)
1v12/	paab00/pp)
1v52/	paac00/pi)
3y11/	uaad00/up)
4j12/bs14/	eaacaa/ette)
5e11/	vaae00/vta0000000000000)
8o11/	aaad00/aarv0t)
8y11/	uaae00/up)

OUTPUT OF PASS 1

```
paac00/pi)aaaa00/aarv0t)haaa00/hm)uaab00/us)naab00/nt3s0n00u0)paaa00/pc)uaab00/us)uaae00/
up)uaaa00/up)uaac00/up)naab00/nt3s0n00u0)paaa00/pc)eaacaa/ette)vaabaa/vi00000000000000)eaa
b00/ette)paab00/pp)naaa00/nm1spm0000)naad00/nc3s0n0000)naacaa/ne3p0m0000)vaae00/vta0000
000000000)eaaa00/ept0)caaaaa/vi00000000000000)nd3s0n0000)haaa00/hm)aaacaa/ajr00t)baaa00/i
aaan)jc)vaaaaa/vta0002000000000)aaab00/adcr0t)aaad00/aarv0t)caaaaa/vi00000000000000)nd3s0n
0000)haaa00/hm)naaeaa/nd3s0ny000)paaa00/pc)vaacaa/vta0000000000000)raaa00/rs)aaaa00/aarv
0t)haaa00/hm)uaab00/us)naab00/nt3s0n00u0)haaa00/hm)vaadaa/vi00000000000000)paab00/pp)
```

DICTIONARY II (IN LINGUISTIC WRITING FORMAT)

Linguistic Pass 1A

1. PC → KCI + PC + KCT
2. PI → PI + KPI

3. $\text{PP} \rightarrow \text{KPT} + \text{PP} + \text{KPI}$

Linguistic Pass 1D

- $\text{IAAAN/JC}_1 + V \sim {}^{16}_{2} + \$_3 + \text{UP TO NEXT} + V \sim \overset{11}{0} \sim {}^{16}/N \sim {}^{10}_{4} \rightarrow \text{JC}_1 + V \sim {}^{16}_{2} + \$_3$
  $+ \text{UP TO NEXT} + V \sim \overset{11}{0} \sim {}^{16}/N \sim {}^{10}_{4}$

Linquistic Pass 1F

- $\text{ADCRXX}_1 + \text{AXRXXX}_2 \rightarrow \quad \text{AXCXXX}_2$

Linguistic Pass 1G

1. $\text{ETIE}_1 + V \sim {}^{16}_{2} + \text{ETTE}_3 \rightarrow \text{KVI} + \text{VXXXXXXXXXXXXXXXEX}_2 + \text{KVT}$
2. $V \sim {}^{16}_{2} + \text{EPT0}_3 \rightarrow \text{KWI} + \text{VXXXXXXXXXXXXP000}_2 + \text{KWT}$

Linguistic Pass 1H

1. $\text{AAXXXX}_1 + \text{HM}_2 \rightarrow \text{AJXXXX}_1 + \text{HM}_2$
2. $\text{AAXXXX} + N \sim {}^{10} \rightarrow \text{AJXXXX} + N \sim {}^{10}$
3. $\text{HM}_1 + \text{AXXXXX}_2 \rightarrow \text{HM}_1 + \text{ND3S0N000A}_2$
4. $\text{HM}_1 + V \sim {}^{16}_{2} \rightarrow \text{HM}_1 + \text{ND3S0N000V}_2$
5. $\begin{bmatrix}\text{KCT}\\ \text{KPI}\end{bmatrix}_1 + \$\overline{\text{VTIC} \sim {}^{16}_{2}} + \text{UP TO NEXT}_3 + \text{AJXXXX}_4 + \text{HM}_5 + \text{UX}_6 + N \sim {}^{10}_{7}$
   $\rightarrow \begin{bmatrix}\text{KCT}\\ \text{KPI}\end{bmatrix}_1 + \$\overline{\text{VTIC} \sim {}^{16}_{2}} + \text{UP TO NEXT}_3 + \text{UX}_6 + \text{AJXXXX}_4 + N \sim {}^{10}_{7}$
6. $\text{US}_1 + \text{UP}_2 + \text{UP}_3 + \text{UP}_4 + \text{NT3S0N00U0}_5 \rightarrow \text{NT3S0N30U0}_5 + \underline{\text{of} + \text{UA}_1 + \text{UA}_2 + \text{UA}_3 + \text{UA}_4}$
7. $\text{VTA} \sim {}^{16}_{1} + \begin{pmatrix}\text{KVT}\\ \text{KWT}\end{pmatrix}_2 + \$\begin{bmatrix}\overline{\underline{N \sim {}^{10}}}\\ \underline{V \sim {}^{16}}\\ \text{KXX}\end{bmatrix}_3 + \text{UP TO NEXT}_4 + \text{VI} \sim {}^{16}/N \sim {}^{10}_{5} \rightarrow \text{VTA} \sim {}^{16}_{1} + \begin{pmatrix}\text{KVT}\\ \text{KWT}\end{pmatrix}_2$
   $+ \$\begin{bmatrix}\overline{\underline{N \sim {}^{10}}}\\ \underline{V \sim {}^{16}}\\ \text{KXX}\end{bmatrix}_3 + \text{UP TO NEXT}_4 + N \sim {}^{10}_{5}$

Linguistic Pass 2A

1. $\text{AJXXX}_1 + N \sim {}^{10}_{2} \rightarrow \underline{\text{AJXXXX}_1} + \text{NXXXXX1XXX}_2$
2. $\text{RS}_1 + \text{US}_2 + \text{AJXXXX}_3 + N \sim {}^{10}_{4} \rightarrow \underline{\text{RS}_1 + \text{AJXXX}_3} + \text{NXXSXX1XXX}_4$
3. $\text{US}_1 + \text{AJXXXX}_2 + N \sim {}^{10}_{3} \rightarrow \underline{\text{a} + \text{AJXXXX}_2} + \text{NXXSXX1XXX}_3$
4. $\overline{\$\text{VTBW} \sim {}^{16} + \begin{pmatrix}\text{KVT}\\ \text{KWT}\end{pmatrix}_1} + \text{NMXXXX0XXX}_2 + \text{NCXXXX0XXX}_3 \rightarrow \overline{\$\text{VTBW} \sim {}^{16} + \begin{pmatrix}\text{KVT}\\ \text{KWT}\end{pmatrix}_1}$
   $+ \underline{\text{NMXPXXXPXX}_2} + \text{NXXXXX1XXX}_3$

5. $\overline{\$\mathrm{VTBW} \sim {}^{16} + \binom{\mathrm{KVT}}{\mathrm{KWT}}_1} + \mathrm{NXXXXX1XXX}_2 + \mathrm{N} \sim {}^{10} \rightarrow \overline{\$\mathrm{VTBW} \sim {}^{16} + \binom{\mathrm{KVT}}{\mathrm{KWT}}_1}\ \underline{+ \text{ of } + \mathrm{NXXXXXXOXX}_2}$

Linguistic Pass 2B

- $\begin{bmatrix}\mathrm{KWT}\\ \mathrm{KVT}\end{bmatrix}_1 + \mathrm{N} \sim {}^{10}_{2} + \mathrm{HM}_3 + \mathrm{N} \sim {}^{10}_{4} + \mathrm{JC}_5 + \mathrm{V} \sim {}^{16}_{6} + \$\overline{\mathrm{KXX}}_7 + \text{UP TO NEXT} + \mathrm{HM}_8$

$\rightarrow \begin{bmatrix}\mathrm{KWT}\\ \mathrm{KVT}\end{bmatrix}_1 + \mathrm{NXXXXX2XXX}_4 + \underline{\text{of} + \mathrm{NXXXXXX0XX}_2} + \mathrm{JC}_5 + \mathrm{V} \sim {}^{16}_{6} + \$\overline{\mathrm{KXX}}_7 + \text{UP TO NEXT}$

$+ \mathrm{HM}_8$

Linguistic Pass 4

1. $\mathrm{KPI} + \mathrm{VTA} \sim {}^{16}_{1} + \mathrm{N} \sim {}^{10}_{2} + \mathrm{HM} + \mathrm{ND} \sim {}^{10}_{3} + \mathrm{KRT} \rightarrow \mathrm{NDXXXXRXXX}_3 + \underline{\text{of}}$
$\underline{+ \mathrm{VTAXXXXXXXXXXP000}_1 + \mathrm{NXXXXXXOXX}_2}$
2. $\mathrm{KXX} + \mathrm{N} \sim {}^{10}_{1} + \mathrm{HM} + \mathrm{N} \sim {}^{10}_{2} + \mathrm{KXX} \rightarrow \mathrm{KXX} + \mathrm{NXXXXX2XXX}_2\ \underline{+ \text{ of } + \mathrm{NXXXXXXOXX}_1}$
$+ \mathrm{KXX}$
3. $\mathrm{VTAXXXXXXXXXXP000}_1 + \mathrm{KWT}_2 + \mathrm{ND} \sim {}^{10}_{3} + \mathrm{JC}_4 + \mathrm{VTA} \sim {}^{16}_{5} + \mathrm{N} \sim {}^{10}_{6} + \mathrm{HM}_7 + \mathrm{ND} \sim {}^{10}_{8}$
$+ \mathrm{KCI}_9 \rightarrow \mathrm{VTAXXXXXXXXXXP000}_1 + \mathrm{KWT}_2 + \mathrm{ND} \sim {}^{10}_{3} + \mathrm{JC}_4 + \mathrm{KRI} + \mathrm{VTA} \sim {}^{16}_{5} + \mathrm{N} \sim {}^{10}_{6}$
$+ \mathrm{HM}_7 + \mathrm{N} \sim {}^{10}_{8} + \mathrm{KRT} + \mathrm{KCI}_9$
4. $\mathrm{VTA} \sim {}^{16}_{1} + \mathrm{KWT}_2 + \mathrm{N} \sim {}^{10}_{3} + \mathrm{JC}_4 + \mathrm{N} \sim {}^{10}_{5} + \mathrm{KCI}_6 + \mathrm{PC} + \mathrm{KCT} + \mathrm{V} \sim {}^{16}_{7} \rightarrow \mathrm{VTA} \sim {}^{16}_{1}$
$+ \mathrm{KWT}_2 + \mathrm{N} \sim {}^{10}_{3} + \mathrm{JC}_4 + \mathrm{N} \sim {}^{10}_{5} + \mathrm{KCI}_6 + \mathrm{PC} + \mathrm{KCT} + \mathrm{KVI} + \mathrm{V} \sim {}^{16}_{7} + \mathrm{KVT}$

Linguistic Pass 5

1. $\mathrm{KPI} + \mathrm{NE3PXXXXXX} + \$\overset{P}{\overline{\mathrm{VXXXXXXXXXXX000XX}}} + \text{UP TO NEXT} + \mathrm{KVI}$
$+ \mathrm{VTAXXXXXXXXX000XX} + \mathrm{KVT} \rightarrow \mathrm{KPI} + \mathrm{NE3PXXXSXX} + \$\overset{P}{\overline{\mathrm{VXXXXXXXXXXX000XX}}}$
$+ \text{UP TO NEXT} + \mathrm{KVI} + \mathrm{VTAXXXXXXXXXV3PXX} + \mathrm{KVT}$
2. $\mathrm{KPI} + \mathrm{NT3SXXXXXX} + \$\overset{P}{\overline{\mathrm{V} \sim {}^{16}}} + \text{UP TO NEXT} + \mathrm{KVI} + \mathrm{VIXXXXXXXXXX000XX}$
$+ \mathrm{KVT} \rightarrow \mathrm{KPI} + \mathrm{NT3SXXXSXX} + \$\overset{P}{\overline{\mathrm{V} \sim {}^{16}}} + \text{UP TO NEXT} + \mathrm{KVI} + \mathrm{VIXXXXXXXXXXV3SXX}$
$+ \mathrm{KVT}$

OUTPUT OF PASS 2

```
paac00/pi)kpi(zazzzz/z)aaaa00/ajrv0t)naab00/nt3s0n10u0)kci(paaa00/pc)kct(naab00/nt3s0n30u0)z
ofzzz/z)uaab00/ua)uaae00/ua)uaaa00/ua)uaac00/ua)kci(paaa00/pc)kct(kvi(vaabaa/vi00000000003s
e0)kvt(kpt(paab00/pp)kpi(naacaa/ne3p0m2s00)zofzzz/z)naaa00/nm1ppm0p00)naad00/nc3s0n0o00)k
wi(vaae00/vta000000000p000)kwt(aaacaa/nd3s0n2o0a)zofzzz/z)caaaaa/nd3s0n0o00)baaa00/jc)naae
```

```
aa/nd3s0nro00)zofzzz/z)vaaaaa/vta000200000p000)aaad00/ajcv0t)caaaaa/nd3s0n0o00)kci(paaa00/
pc)kct(kvi(vaacaa/vta0000000003p0a)kvt(vaadaa/nd3s0n2o0v)zofzzz/z)raaa00/rs)aaaa00/ajrv0t)na
ab00/nt3s0n1o00)kpt(paab00/pp)kpi(
```

DICTIONARY III

Pseudo-Chicodes and English Grammatic Tags	English Translation Stem and English Inflection Table Tags
aaaa00/a-----)	new0ael
aaab00/adca--)	
aaab00/adrv--)	even more0
aaacaa/a-----)	joyous0aml
aaacaa/n---------)	joy0ni0b
aaad00/a-----)	great0ael
baaa00/ia1--)	with0
baaa00/ia2--)	as0
baaa00/ia3--)	from0
baaa00/j-)	and0
caaaaa/a-----)	victorious0aml
caaaaa/n---------)	victor0ny0a
caaaaa/v----------------)	b victorious0vb01
eaaa00/e---)	
eaab00/e---)	
eaac00/e---)	
haaa00/hm)	
naaa00/n--p---n--)	ours0
naaa00/n--p---o--)	us0
naaa00/n--p---p--)	our0
naaa00/n--p---s--)	we0
naaa00/n--s---n--)	mine0
naaa00/n--s---o--)	me0
naaa00/n--s---p--)	my0
naaa00/n--s---s--)	I0
naab00/n---------)	year0ns0a
naacaa/n---------)	people0np0b

naad00/n---------)	countr0ny0a
naaeaa/a-----)	confident0aml
naaeaa/n---------)	confidence0ni0b
paaa00/p-)	,0
paab00/p-)	.00
paac00/p-)	00000
raaa00/rp)	these0
raaa00/rs)	this0
uaaa00/ua)	60
uaaa00/ui)	six0
uaaa00/uq)	sixth0
uaaa00/ut)	sixty0
uaab00/ua)	10
uaab00/ui	one0
uaab00/uq)	first0
uaab00/ut)	ten0
uaac00/ua)	50
uaac00/ui)	five0
uaac00/uq)	fifth0
uaac00/ut)	fifty0
uaad00/ua)	40
uaad00/ui)	four0
uaad00/uq)	fourth0
uaad00/ut)	forty0
uaae00/ua)	90
uaae00/ui)	nine0
uaae00/uq)	ninth0
uaae00/ut)	ninety0
vaaaaa/n---------)	striving for0ni0b
vaaaaa/v---------------)	striv for0vv01v5strove for0v6striven for0
vaabaa/n---------)	arrival0ni0b
vaabaa/v---------------)	arriv0vve0
vaacaa/n---------)	welcome0ns0a
vaacaa/v---------------)	welcom0vve0
vaadaa/n---------)	arrival0ns0a

```
vaadaa/v---------------)    arriv0vve0
vaae00/n---------)          bosom0ns0a
vaae00/v---------------)    embrac0vve0
zazzzz/z)                   a0
zofzzz/z)                   of0
```

OUTPUT OF PASS 3

```
00000a0new0aelyear0ns0a,0year0ns0aof010906050,0arriv0vve0.00people0np0bof0our0countr0ny0a
embrac0vve0joy0ni0bof0victor0ny0aand0confidence0ni0bof0striv for0vv01v5strove for0v6striven
for0great0aelvictor0ny0a,0welcom0vve0arrival0ns0aof0this0new0aelyear0ns0a.00
```

FINAL OUTPUT (AFTER INFLECTION PROCESS)

A new year , year of 1 9 6 5 , has arrived . People of our country embracing joy of victory and confidence of striving for greater victory , welcome arrival of this new year .

Appendix C

VERB COMPONENTS IN ENGLISH OUTPUTS

The form a verb will take in the final English output is determined by positions 12 through 16 of the verb tag word. Tables C-2 through C-5 are designed as a guide to programming to show which verb forms are to be combined for each possible configuration of verb tags 12 through 16. The code letters shown in Table C-1 are utilized.

POSITIVE VERBS

A verb may take three person forms (first, second, and third), two numbers (singular and plural), six tenses (present, present perfect, past, past perfect, future, and future perfect), two forms (regular and progressive), and two voices (active and passive). Thus 3 × 2 × 6 × 2 × 2 yields a total of 144 possible forms for each verb (exclusive of negative, infinitive, and postmodal forms given below).

Table C-2 shows the English tag code for each of the possible 144 forms (tags are lacking for all progressive tenses but the present) and the form of the auxiliary and verb stem to be used in each case. Examples are given using the verb "to see" and the verb "to be."

NEGATIVE VERBS

There are also 144 negative verb forms, as shown in Table C-3.

INFINITIVE FORMS

Table C-4 shows all possible infinitive forms, positive and negative, for all verbs and the verb "to be" both by itself and in compounds like "to be stupid."

POSTMODAL VERBS

The possible combinations for regular verbs and for verbs of the "to be (stupid)" type may be formed according to the scheme shown in Table C-5.

Table C-1 — Code Letters

Tag Number	Letter	Meaning
12	V	Main verb
	Y	Negative
13	1	First person
	2	Second person
	3	Third person
	0	Does not have
	P	Present participle
	D	Past participle
	I	Infinitive
	G	Auxiliary
14	S	Singular
	P	Plural
	0	Does not have
15	P	Present
	E	Present perfect
	A	Past
	S	Past perfect
	F	Future
	U	Future perfect
	R	Present progressive
16	A	Active
	P	Passive

Table C-2 — English Tag Codes for All Positive Verb Forms

English Tags 12 to 16	Verb Form		"To Be" Form		Examples		
	Auxiliary	Stem	Auxiliary	Stem	Auxiliary	To See	To Be
V1SPA		V1		B1		see	am
V2SPA		V1		B2		see	are
V3SPA		V2		B3		sees	is
V1PPA		V1		B2		see	are
V2PPA		V1		B2		see	are
V3PPA		V1		B2		see	are
V1SEA	A1	V6	A1	B9	have	seen	been
V2SEA	A1	V6	A1	B9	have	seen	been
V3SEA	A2	V6	A2	B9	has	seen	been
V1PEA	A1	V6	A1	B9	have	seen	been
V2PEA	A1	V6	A1	B9	have	seen	been
V3PEA	A1	V6	A1	B9	have	seen	been
V1SAA		V5		B4		saw	was
V2SAA		V5		B5		saw	were
V3SAA		V5		B4		saw	was
V1PAA		V5		B5		saw	were
V2PAA		V5		B5		saw	were
V3PAA		V5		B5		saw	were
V1SSA	A3	V6	A3	B9	had	seen	been
V2SSA	A3	V6	A3	B9	had	seen	been
V3SSA	A3	V6	A3	B9	had	seen	been
V1PSA	A3	V6	A3	B9	had	seen	been
V2PSA	A3	V6	A3	B9	had	seen	been
V3PSA	A3	V6	A3	B9	had	seen	been
V1SFA	A4	V1	A4	B7	will	see	be
V2SFA	A4	V1	A4	B7	will	see	be
V3SFA	A4	V1	A4	B7	will	see	be
V1PFA	A4	V1	A4	B7	will	see	be
V2PFA	A4	V1	A4	B7	will	see	be
V3PFA	A4	V1	A4	B7	will	see	be

Table C-2 — English Tag Codes for All Positive Verb Forms (Cont.)

English Tags 12 to 16	Verb Form		"To Be" Form		Examples		
	Auxiliary	Stem	Auxiliary	Stem	Auxiliary	To See	To Be
V1SUA	A5	V6	A5	B9	will have	seen	been
V2SUA	A5	V6	A5	B9	will have	seen	been
V3SUA	A5	V6	A5	B9	will have	seen	been
V1PUA	A5	V6	A5	B9	will have	seen	been
V2PUA	A5	V6	A5	B9	will have	seen	been
V3PUA	A5	V6	A5	B9	will have	seen	been
V1SRA	A6	V3	A6	B8	am	seeing	being
V2SRA	A7	V3	A7	B8	are	seeing	being
V3SRA	A8	V3	A8	B8	is	seeing	being
V1PRA	A7	V3	A7	B8	are	seeing	being
V2PRA	A7	V3	A7	B8	are	seeing	being
V3PRA	A7	V3	A7	B8	are	seeing	being
V1S_A	A9	V3	A9	B8	have been	seeing	being
V2S_A	A9	V3	A9	B8	have been	seeing	being
V3S_A	A10	V3	A10	B8	has been	seeing	being
V1P_A	A9	V3	A9	B8	have been	seeing	being
V2P_A	A9	V3	A9	B8	have been	seeing	being
V3P_A	A9	V3	A9	B8	have been	seeing	being
V1S_A	A11	V3	A11	B8	was	seeing	being
V2S_A	A12	V3	A12	B8	were	seeing	being
V3S_A	A11	V3	A11	B8	was	seeing	being
V1P_A	A12	V3	A12	B8	were	seeing	being
V2P_A	A12	V3	A12	B8	were	seeing	being
V3P_A	A12	V3	A12	B8	were	seeing	being
V1S_A	A13	V3	A13	B8	had been	seeing	being
V2S_A	A13	V3	A13	B8	had been	seeing	being
V3S_A	A13	V3	A13	B8	had been	seeing	being
V1P_A	A13	V3	A13	B8	had been	seeing	being
V2P_A	A13	V3	A13	B8	had been	seeing	being
V3P_A	A13	V3	A13	B8	had been	seeing	being

Table C-2 — English Tag Code for All Positive Verb Forms (Cont.)

English Tags	Verb Form		"To Be" Form		Examples		
12 to 16	Auxiliary	Stem	Auxiliary	Stem	Auxiliary	To See	To Be
V1S_A	A14	V3	A14	B8	will be	seeing	being
V2S_A	A14	V3	A14	B8	will be	seeing	being
V3S_A	A14	V3	A14	B8	will be	seeing	being
V1P_A	A14	V3	A14	B8	will be	seeing	being
V2P_A	A14	V3	A14	B8	will be	seeing	being
V3P_A	A14	V3	A14	B8	will be	seeing	being
V1S_A	A15	V3	A15	B8	will have been	seeing	being
V2S_A	A15	V3	A15	B8	will have been	seeing	being
V3S_A	A15	V3	A15	B8	will have been	seeing	being
V1P_A	A15	V3	A15	B8	will have been	seeing	being
V2P_A	A15	V3	A15	B8	will have been	seeing	being
V3P_A	A15	V3	A15	B8	will have been	seeing	being
V1SPP	A6	V6			am	seen	
V2SPP	A7	V6			are	seen	
V3SPP	A8	V6			is	seen	
V1PPP	A7	V6			are	seen	
V2PPP	A7	A6			are	seen	
V3PPP	A7	V6			are	seen	
V1SEP	A9	V6			have been	seen	
V2SEP	A9	V6			have been	seen	
V3SEP	A10	V6			has been	seen	
V1PEP	A9	V6			have been	seen	
V2PEP	A9	V6			have been	seen	
V3PEP	A9	V6			have been	seen	
V1SAP	A11	V6			was	seen	
V2SAP	A12	V6			were	seen	
V3SAP	A11	V6			was	seen	
V1PAP	A12	V6			were	seen	
V2PAP	A12	V6			were	seen	
V3PAP	A12	V6			were	seen	

Table C-2 — English Tag Code for All Positive Verb Forms (Cont.)

English Tags 12 to 16	Verb Form		"To Be" Form		Examples		
	Auxiliary	Stem	Auxiliary	Stem	Auxiliary	To See	To Be
V1SSP	A13	V6			had been	seen	
V2SSP	A13	V6			had been	seen	
V3SSP	A13	V6			had been	seen	
V1PSP	A13	V6			had been		
V2PSP	A13	V6			had been	seen	
V3PSP	A13	V6			had been	seen	
V1SFP	A14	V6			will be	seen	
V2SFP	A14	V6			will be	seen	
V3SFP	A14	V6			will be	seen	
V1PFP	A14	V6			will be	seen	
V2PFP	A14	V6			will be	seen	
V3PFP	A14	V6			will be	seen	
V1SUP	A15	V6			will have been	seen	
V2SUP	A15	V6			will have been	seen	
V3SUP	A15	V6			will have been	seen	
V1PUP	A15	V6			will have been	seen	
V2PUP	A15	V6			will have been	seen	
V3PUP	A15	V6			will have been	seen	
V1SRP	A16	V6			am being	seen	
V2SRP	A17	V6			are being	seen	
V3SRP	A18	V6			is being	seen	
V1PRP	A17	V6			are being	seen	
V2PRP	A17	V6			are being	seen	
V3PRP	A17	V6			are being	seen	
V1S_P	–	V6			(have been being)	seen	
V2S_P	–	V6			(have been being)	seen	
V3S_P	–	V6			(has been being)	seen	
V1P_P	–	V6			(have been being)	seen	
V2P_P	–	V6			(have been being)	seen	
V3P_P	–	V6			(have been being)	seen	

Table C-2 — English Tag Code for All Positive Verb Forms (Cont.)

English Tags 12 to 16	Verb Form		"To Be" Form		Examples		
	Auxiliary	Stem	Auxiliary	Stem	Auxiliary	To See	To Be
V1S_P	–	V6			(was being)	seen	
V2S_P	–	V6			(were being)	seen	
V3S_P	–	V6			(was being)	seen	
V1P_P	–	V6			(were being)	seen	
V2P_P	–	V6			(were being)	seen	
V3P_P	–	V6			(were being)	seen	
V1S_P	–	V6			(had been being)	seen	
V2S_P	–	V6			(had been being)	seen	
V3S_P	–	V6			(had been being)	seen	
V1P_P	–	V6			(had been being)	seen	
V2P_P	–	V6			(had been being)	seen	
V3P_P	–	V6			(had been being)	seen	
V1S_P	–	V6			(will be being)	seen	
V2S_P	–	V6			(will be being)	seen	
V3S_P	–	V6			(will be being)	seen	
V1P_P	–	V6			(will be being)	seen	
V2P_P	–	V6			(will be being)	seen	
V3P_P	–	V6			(will be being)	seen	
V1S_P	–	V6			(will have been being)	seen	
V2S_P	–	V6			(will have been being)	seen	
V3S_P	–	V6			(will have been being)	seen	
V1P_P	–	V6			(will have been being)	seen	
V2P_P	–	V6			(will have been being)	seen	
V3P_P	–	V6			(will have been being)	seen	

Table C-3 — English Tag Codes for All Negative Verb Forms

English Tags 12 to 16	Verb Form		"To Be" Form		Examples	
	Auxiliary	Stem	Auxiliary	Stem	To See	To Be
Y1SPA	D1	V1	D9	–	do not see	am not + 0
Y2SPA	D1	V1	D10	–	do not see	are not + 0
Y3SPA	D2	V1	D11	–	does not see	is not + 0
Y1PPA	D1	V1	D10	–	do not see	are not + 0
Y2PPA	D1	V1	D10	–	do not see	are not + 0
Y3PPA	D1	V1	D10	–	do not see	are not + 0
Y1SEA	D3	V6	D3	B9	have not seen	have not been
Y2SEA	D3	V6	D3	B9	have not seen	have not been
Y3SEA	D4	V6	D4	B9	has not seen	has not been
Y1PEA	D3	V6	D3	B9	have not seen	have not been
Y2PEA	D3	V6	D3	B9	have not seen	have not been
Y3PEA	D3	V6	D3	B9	have not seen	have not been
Y1SAA	D5	V1	D14	–	did not see	was not + 0
Y2SAA	D5	V1	D15	–	did not see	were not + 0
Y3SAA	D5	V1	D14	–	did not see	was not + 0
Y1PAA	D5	V1	D15	–	did not see	were not + 0
Y2PAA	D5	V1	D15	–	did not see	were not + 0
Y3PAA	D5	V1	D15	–	did not see	were not + 0
Y1SSA	D6	V6	D6	B9	had not seen	had not been
Y2SSA	D6	V6	D6	B9	had not seen	had not been
Y3SSA	D6	V6	D6	B9	had not seen	had not been
Y1PSA	D6	V6	D6	B9	had not seen	had not been
Y2PSA	D6	V6	D6	B9	had not seen	had not been
Y3PSA	D6	V6	D6	B9	had not seen	had not been
Y1SFA	D7	V1	D7	B7	will not see	will not be
Y2SFA	D7	V1	D7	B7	will not see	will not be
Y3SFA	D7	V1	D7	B7	will not see	will not be
Y1PFA	D7	V1	D7	B7	will not see	will not be
Y2PFA	D7	V1	D7	B7	will not see	will not be
Y3PFA	D7	V1	D7	B7	will not see	will not be

Table C-3 — English Tag Codes for All Negative Verb Forms (Cont.)

English Tags 12 to 16	Verb Form		"To Be" Form		Examples	
	Auxiliary	Stem	Auxiliary	Stem	To See	To Be
Y1SUA	D8	V6	D8	B9	will not have seen	will not have been
Y2SUA	D8	V6	D8	B9	will not have seen	will not have been
Y3SUA	D8	V6	D8	B9	will not have seen	will not have been
Y1PUA	D8	V6	D8	B9	will not have seen	will not have been
Y2PUA	D8	V6	D8	B9	will not have seen	will not have been
Y3PUA	D8	V6	D8	B9	will not have seen	will not have been
Y1SRA	D9	V3	D9	B8	am not seeing	am not being
Y2SRA	D10	V3	D10	B8	are not seeing	are not being
Y3SRA	D11	V3	D11	B8	is not seeing	is not being
Y1PRA	D10	V3	D10	B8	are not seeing	are not being
Y2PRA	D10	V3	D10	B8	are not seeing	are not being
Y3PRA	D10	V3	D10	B8	are not seeing	are not being
Y1S_A	D12	V3	D12	B8	have not been seeing	have not been being
Y2S_A	D12	V3	D12	B8	have not been seeing	have not been being
Y3S_A	D13	V3	D13	B8	has not been seeing	has not been being
Y1P_A	D12	V3	D12	B8	have not been seeing	have not been being
Y2P_A	D12	V3	D12	B8	have not been seeing	have not been being
Y3P_A	D12	V3	D12	B8	have not been seeing	have not been being
Y1S_A	D14	V3	D14	B8	was not seeing	was not being
Y2S_A	D15	V3	D15	B8	were not seeing	were not being
Y3S_A	D14	V3	D14	B8	was not seeing	was not being
Y1P_A	D15	V3	D15	B8	were not seeing	were not being
Y2P_A	D15	V3	D15	B8	were not seeing	were not being
Y3P_A	D15	V3	D15	B8	were not seeing	were not being
Y1S_A	D16	V3	D16	B8	had not been seeing	had not been being
Y2S_A	D16	V3	D16	B8	had not been seeing	had not been being
Y3S_A	D16	V3	D16	B8	had not been seeing	had not been being
Y1P_A	D16	V3	D16	B8	had not been seeing	had not been being
Y2P_A	D16	V3	D16	B8	had not been seeing	had not been being
Y3P_A	D16	V3	D16	D8	had not been seeing	had not been being

Table C-3 — English Tag Codes for All Negative Verb Forms (Cont.)

English Tags 12 to 16	Verb Form		"To Be" Form		Examples	
	Auxiliary	Stem	Auxiliary	Stem	To See	To Be
Y1S_A	D17	V3	D17	B8	will not be seeing	will not be being
Y2S_A	D17	V3	D17	B8	will not be seeing	will not be being
Y3S_A	D17	V3	D17	B8	will not be seeing	will not be being
Y1P_A	D17	V3	D17	B8	will not be seeing	will not be being
Y2P_A	D17	V3	D17	B8	will not be seeing	will not be being
Y3P_A	D17	V3	D17	B8	will not be seeing	will not be being
Y1S_A	D18	V3	D18	B8	will not have been seeing	will not have been being
Y2S_A	D18	V3	D18	B8	will not have been seeing	will not have been being
Y3S_A	D18	V3	D18	B8	will not have been seeing	will not have been being
Y1P_A	D18	V3	D18	B8	will not have been seeing	will not have been being
Y2P_A	D18	V3	D18	B8	will not have been seeing	will not have been being
Y3P_A	D18	V3	D18	B8	will not have been seeing	will not have been being
Y1SPP	D9	V6			am not seen	
Y2SPP	D10	V6			are not seen	
Y3SPP	D11	V6			is not seen	
Y1PPP	D10	V6			are not seen	
Y2PPP	D10	V6			are not seen	
Y3PPP	D10	V6			are not seen	
Y1SEP	D12	V6			have not been seen	
Y2SEP	D12	V6			have not been seen	
Y3SEP	D13	V6			has not been seen	
Y1PEP	D12	V6			have not been seen	
Y2PEP	D12	V6			have not been seen	
Y3PEP	D12	V6			have not been seen	
Y1SAP	D14	V6			was not seen	
Y2SAP	D15	V6			were not seen	
Y3SAP	D14	V6			was not seen	
Y1PAP	D15	V6			were not seen	
Y2PAP	D15	V6			were not seen	
Y3PAP	D15	V6			were not seen	

Table C-3 — English Tag Codes for All Negative Verb Forms (Cont.)

English Tags 12 to 16	Verb Form		"To Be" Form		Examples	
	Auxiliary	Stem	Auxiliary	Stem	To See	To Be
Y1SSP	D16	V6			had not been seen	
Y2SSP	D16	V6			had not been seen	
Y3SSP	D16	V6			had not been seen	
Y1PSP	D16	V6			had not been seen	
Y2PSP	D16	V6			had not been seen	
Y3PSP	D16	V6			had not been seen	
Y1SFP	D17	V6			will not be seen	
Y2SFP	D17	V6			will not be seen	
Y3SFP	D17	V6			will not be seen	
Y1PFP	D17	V6			will not be seen	
Y2PFP	D17	V6			will not be seen	
Y3PFP	D17	V6			will not be seen	
Y1SUP	D18	V6			will not have been seen	
Y2SUP	D18	V6			will not have been seen	
Y3SUP	D18	V6			will not have been seen	
Y1PUP	D18	V6			will not have been seen	
Y2PUP	D18	V6			will not have been seen	
Y3PUP	D18	V6			will not have been seen	
Y1SRP	D19	V6			am not being seen	
Y2SRP	D20	V6			are not being seen	
Y3SRP	D21	V6			is not being seen	
Y1PRP	D20	V6			are not being seen	
Y2PRP	D20	V6			are not being seen	
Y3PRP	D20	V6			are not being seen	
Y1S_P	D_	V6			(have not been being) seen	
Y2S_P	D_	V6			(have not been being) seen	
Y3S_P	D_	V6			(has not been being) seen	
Y1P_P	D_	V6			(have not been being) seen	
Y2P_P	D_	V6			(have not been being) seen	
Y3P_P	D_	V6			(have not been being) seen	

Table C-3 — English Tag Codes for All Negative Verb Forms (Cont.)

English Tags 12 to 16	Verb Form		"To Be" Form		Examples	
	Auxiliary	Stem	Auxiliary	Stem	To See	To Be
Y1S_P	D_	V6			(was not being) seen	
Y2S_P	D_	V6			(were not being) seen	
Y3S_P	D_	V6			(was not being) seen	
Y1P_P	D_	V6			(were not being) seen	
Y2P_P	D_	V6			(were not being) seen	
Y3P_P	D_	V6			(were not being) seen	
Y1S_P	D_	V6			(had not been being) seen	
Y2S_P	D_	V6			(had not been being) seen	
Y3S_P	D_	V6			(had not been being) seen	
Y1P_P	D_	V6			(had not been being) seen	
Y2P_P	D_	V6			(had not been being) seen	
Y3P_P	D_	V6			(had not been being) seen	
Y1S_P	D_	V6			(will not be being) seen	
Y2S_P	D_	V6			(will not be being) seen	
Y3S_P	D_	V6			(will not be being) seen	
Y1P_P	D_	V6			(will not be being) seen	
Y2P_P	D_	V6			(will not be being) seen	
Y3P_P	D_	V6			(will not be being) seen	
Y1S_P	D_	V6			(will not have been being) seen	
Y2S_P	D_	V6			(will not have been being) seen	
Y3S_P	D_	V6			(will not have been being) seen	
Y1P_P	D_	V6			(will not have been being) seen	
Y2P_P	D_	V6			(will not have been being) seen	
Y3P_P	D_	V6			(will not have been being) seen	

Table C-4 — Infinitive Forms for All Verbs

English Tags 12 to 16	Verb Form		"To Be" Form		Examples		
	Infinitive	Stem	Infinitive	Stem	Auxiliary	To See	To Be
-IXPA		V4		B6		to see	to be
-IXEA	I2	V6	I2	B9	to have	seen	been
-IXRA	I3	V3	I3	B8	to be	seeing	being
-IX_A		V3		B8	(to have been)	seeing	being
-IXPP	I4	V6			to be	seen	
-IXEP	I5	V6			to have been	seen	
-IXRP		V6			(to be being)	seen	
-IX_P		V6			(to have been being)	seen	
YIXPA	I6	V1	I6	B7	to not	see	be
YIXEA	I7	V6	I7	B9	to not have	seen	been
YIXRA	I8	V3	I8	B8	to not be	seeing	being
YIX_A		V3		B8	(to not have been)	seeing	being
YIXPP	I9	V6			to not be	seen	
YIXEP	I10	V6			to not have been	seen	
YIXRP		V6			(to not be being)	seen	
YIX_P		V6			(to not have been being)	seen	

Table C-5 — Possible Verb Combinations

English Tags 12 to 16	Verb Form G	Verb Form Stem	"To Be" Form	Modal	Examples To Finish	Examples To Be	Examples (Able)
0G00A		V1	G1	must	finish	be	(able)
0G00P	G1	V6	G1	must	be finished	be	(able)
0G0EA	G3	V6	G4	must	have finished	have been	(able)
0G0EP	G4	V6	G4	must	have been finished	have been	(able)
0G0RA	G1	V3	G2	must	be finishing	be being	(able)
0G0RP	G2	V6	G2	must	be being finished	be being	(able)
0G0FA		V1	G1	must	finish	be	(able)
0G0FP	G1	V6	G1	must	be finished	be	(able)
0G0AA	G3	V6	G4	should	have finished	have been	(able)
0G0AP	G4	V6	G4	should	have been finished	have been	(able)
YG00A	G5	V1	G6	must	not finish	not be	(able)
YG00P	G6	V6	G6	must	not be finished	not be	(able)
YG0EA	G8	V6	G9	must	not have finished	not have been	(able)
YG0EP	G9	V6	G9	must	not have been finished	not have been	(able)
YG0RA	G6	V3	G7	must	not be finishing	not be being	(able)
YG0RP	G7	V6	G7	must	not be being finished	not be being	(able)
YG0FA	G5	V1	G6	must	not finish	not be	(able)
YG0FP	G6	V6	G6	must	not be finished	not be	(able)
YG0AA	G8	V6	G9	should	not have finished	not have been	(able)
YG0AP	G9	V6	G9	should	not have been finished	not have been	(able)

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Itek Corporation
10 Maguire Road
Lexington 73, Massachusetts

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP
N/A

3. REPORT TITLE
Linguistic Studies for Chinese to English Machine Translation

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Final 1 July 64 - 30 June 65

5. AUTHOR(S) *(Last name, first name, initial)*
Mrs. Lee, Theresa et al.

6. REPORT DATE
June 1966

7a. TOTAL NO. OF PAGES
196

7b. NO. OF REFS
0

8a. CONTRACT OR GRANT NO.
AF30(602)-3494

b. PROJECT NO.
4599

c.
62405454

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
RADC-TR-65-443

10. AVAILABILITY/LIMITATION NOTICES
This document is subject to special export controls and each transmittal to foreig governments or foreign nationals may be made only with prior approval of RADC (EMLI), GAFB, N.Y. 13440.

11. SUPPLEMENTARY NOTES
Applied linguistic research for Chinese-English machine aided translation

12. SPONSORING MILITARY ACTIVITY
Rome Air Development Center (EMIIH)
Griffiss AFB NY 13440.

13. ABSTRACT

The linguistic study for a developmental Chinese-English machine aided translation resulted in the design of a basic linguistic processing system based on the Contextual Associative Method (CAM). This technique allows machine aided translation through the use of programmed contextual operations.

The results of this research effort are presented in this report and and include: (1) explanation of the linguistic processing system; (2) morphological and syntactic analyses, and (3) English inflection analysis for Chinese to English machine aided translation. Illustrations showing step by step linguistic processing are included in this report. Recommendations are presented for refinement and further development of the basic linguistic analysis. Three appendices are included: an explanation of symbols for linguistic rules (Appendix A), listings of computer experimentation (Appendix B), and listings of verb components in English output (Appendix C).

DD FORM 1 JAN 64 1473

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Documentation: Machine Translation						
Social Science: Language						

INSTRUCTIONS

1. **ORIGINATING ACTIVITY:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. **REPORT SECURITY CLASSIFICATION:** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. **GROUP:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **REPORT TITLE:** Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. **DESCRIPTIVE NOTES:** If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **AUTHOR(S):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE:** Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. **TOTAL NUMBER OF PAGES:** The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. **NUMBER OF REFERENCES:** Enter the total number of references cited in the report.

8*a*. **CONTRACT OR GRANT NUMBER:** If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. **PROJECT NUMBER:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. **ORIGINATOR'S REPORT NUMBER(S):** Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. **OTHER REPORT NUMBER(S):** If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. **AVAILABILITY/LIMITATION NOTICES:** Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known

11. **SUPPLEMENTARY NOTES:** Use for additional explanatory notes.

12. **SPONSORING MILITARY ACTIVITY:** Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. **ABSTRACT:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional